

UNIVERZITET U BEOGRADU
BIOLOŠKI FAKULTET

Branislava S. Gemović

BIOINFORMATIČKA ANALIZA PROTEINA
UKLJUČENIH U PATOGENEZU
MIJELOIDNIH MALIGNITETA

doktorska disertacija

Beograd, 2015

UNIVERSITY OF BELGRADE
FACULTY OF BIOLOGY

Branislava S. Gemović

**BIOINFORMATICS ANALYSIS OF
PROTEINS INVOLVED IN PATHOGENESIS
OF MYELOID MALIGNANCIES**

Doctoral Dissertation

Belgrade, 2015

Mentori:

dr Dušanka Savić Pavićević

vanredni profesor Biološkog fakulteta Univerziteta u Beogradu

dr Nevena Veljković

naučni savetnik Instituta za nuklearne nauke Vinča Univerziteta u Beogradu

Komisija:

dr Dragomir Marisavljević

redovni profesor Medicinskog fakulteta Univerziteta u Beogradu

dr Katarina Nikolić

docent Farmaceutskog fakulteta Univerziteta u Beogradu

dr Goran Brajušković

vanredni profesor Biološkog fakulteta Univerziteta u Beogradu

Datum odbrane:

Zahvalnica

Ovaj rad je urađen u Centru za multidisciplinarna istraživanja i inženjerstvo Instituta za nuklearne nauke Vinča Univerziteta u Beogradu u okviru projekta Ministarstva prosvete, nauke i tehnološkog razvoja ON173001: „Primena EIIP/ISM bioinformatičke platforme u otkrivanju novih terapijskih targeta i potencijalnih terapijskih molekula“.

Najveću zahvalnost dugujem mojoj mentorki, dr Neveni Veljković. Ona je budno pratila moj istraživački rad od samog početka, usmeravajući ga, a njena podrška, strpljenje i veliko znanje je omogućilo da uspešno dođemo do cilja. Uz pomoć dr Veljković sam odabrala oblast istraživanja, realizovala svaki korak ovog rada i, na kraju, napisala ovu disertaciju. Za to, i još mnogo bitnih sitnica u svakodnevnom zajedničkom radu, hvala.

Želim da izrazim veliku zahvalnost i mojoj mentorki dr Dušanki Savić Pavićević, profesorki Biološkog fakulteta Univerziteta u Beogradu, za komentare i savete prilikom izrade i pisanja disertacije. Ona mi je značajno pomogla da razumem relevantne molekularno biološke koncepte i da rezultate i zaključke svojih istraživanja predstavim na jasan i precizan način.

Zahvaljujem se i dr Dragomiru Marisavljeviću, profesoru Medicinskog fakulteta Univerziteta u Beogradu, i dr Katarini Nikolić, profesorki Farmaceutskog fakulteta Univerziteta u Beogradu, na ljubaznosti, spremnosti na saradnju i produktivnoj komunikaciji u toku istraživanja i pisanja doktorata.

Zahvalnost dugujem i dr Goranu Brajuškoviću, profesoru Biološkog fakulteta Univerziteta u Beogradu, sa kojim sam saradivala i tokom izrade diplomskog rada.

Izrada ovog rada ne bi bila moguća bez pomoći i podrške mojih kolega iz Centra zbog čega sam im neizmerno zahvalna. Posebnu zahvalnost dugujem dr Vladimiru Peroviću, koji je kreirao sve programe koje sam koristila u ovom istraživanju. Zahvalna sam mu i na razgovorima i objašnjenjima matematičkih osnova mnogih metoda, na važnim sugestijama

prilikom izrade ove disertacije, kao i na divnom drugarstvu, i na poslu i van njega. Zahvaljujem se i dr Sanji Glišić na stalnoj spremnosti da pomogne i značajnim savetima i preporukama, kojima je doprinela realizaciji ovog istraživanja. Hvala i Nebojši Škrbiću za tehničku podršku prilikom izrade ove disertacije.

Želim da pomenem i dr Jude Fitzgibbon koji mi je omogućio rad u Centru za Hemato-Onkologiju na Barts and The London School of Medicine and Dentistry, Queen Mary University of London. Mesec dana života i rada u Londonu je za mene bilo neprocenjivo iskustvo. Želim da se zahvalim i dr Stefan Roberts i njegovoj grupi na College of Arts and Sciences, University at Buffalo, The State University of New York, za interesovanje i spremnost da eksperimentalno testiraju neke rezultate dobijene u okviru ove disertacije.

Pomoć, podrška i stalno navijanje mojih roditelja, Slavenke i Slobodana, su mi od neprocenjive važnosti. Ogroman uticaj na moju odluku da se bavim naukom su imali i moja sestra Aleksandra i zet Milutin. Oni su mi pružali stalnu podršku tokom rada na doktoratu i zato im, pored svega ostalog, dugujem veliku zahvalnost. Ovo je mesto i da se zahvalim na svemu mojoj drugoj porodici, mojim prijateljima, Kristini, Maji, Anji, Ani, Goranu i Iloni. Na kraju, ostaje Boban, ali nema dovoljno reči da njemu izrazim svu zahvalnost, zato, eto, samo hvala!

Bioinformatička analiza proteina uključenih u patogenezu mijeloidnih maligniteta

Rezime

Mijeloidni maligniteti su klonalne bolesti ćelija mijeloidne krvne loze, koje su, pre svega, posledica poremećaja samoobnavljanja i diferencijacije hematopoetskih matičnih ćelija. Godišnja incidenca u Evropi je 7,5-8,6/100 000, a prosečna petogodišnja stopa preživljavanja pacijenata, uz primenu standardne terapije, je 37%. S obzirom da maligniteti najčešće nose više od jedne vodeće – ‘*driver*’ mutacije, pri čemu je velika intra- i interkancerska genetička heterogenost, postoji potreba za individualizovanim terapijskim pristupom svakom pacijentu. Personalizovana medicina označava princip postizanja maksimalne terapijske efikasnosti kroz upotrebu ciljanih terapeutika kod biološki okarakterisanih pacijenata. Bioinformatika ima značajnu ulogu u identifikaciji ciljnih molekula za razvoj terapije, racionalnom dizajnu lekova i identifikaciji informativnih biomarkera za prateću dijagnostiku.

Predmet istraživanja u ovom radu bili su geni i njihovi proteinski produkti koji imaju važnu ulogu u patogenezi mijeloidnih maligniteta. WT1 je transkripcioni faktor, koji kontroliše ekspresiju gena uključenih u apoptozu, proliferaciju i diferencijaciju. Ovaj gen ima povećanu ekspresiju u 70-90%, a mutiran je u ~10% akutnih mijeloidnih leukemija (AML) i učestvuje u preko 40 protein-protein interakcija, koje su vremenski i kontekstno zavisne. Mapiranje celokupne mreže WT1 kofaktora doprineće identifikaciji novih molekula i interakcija, koje su potencijalni targeti za ciljanu terapiju AML.

Nedavno otkriven tumor supresor u karcinomu dojke, NISCH, ispoljava efekat preko RAC signalnog puta, koji je ključan u regulaciji hematopoeze i uključen u patogenezu mijeloidnih maligniteta, i smatra se atraktivnim targetom za ciljanu terapiju. Poznavanje

funkcionalnih domena NISCH-a će omogućiti dizajn i optimizaciju jedinjenja sa antikancerskim delovanjem.

Poslednjih desetak godina je utvrđen značaj epigenetičkih regulatora, koji su, pored transkripcionih faktora i signalnih molekula, neophodni za normalno funkcionisanje hematopoetskog sistema. Ovi molekuli imaju tumor supresorski efekat, a mutacije u ASXL1, DNMT3A, EZH2 i TET2 su prisutne u svim tipovima mijeloidnih maligniteta i pretpostavlja se da dovode do nastanka ovih bolesti. Detekcija ovih štetnih mutacija je važan prateći dijagnostik za terapije koji utiču na epigenom.

U ovom radu su primenjeni bioinformatički programi zasnovani na metodi informacionog spektra (ISM) za: i) predviđanje protein-protein interakcija (PPI) WT1, ii) identifikaciju i karakterizaciju funkcionalnih domena proteina NISCH i iii) predviđanje funkcionalnih efekata aminokiselinskih supstitucija (AAS) u epigenetičkim regulatorima ASXL1, DNMT3A, EZH2 i TET2. ISM je metoda za analizu sekvenci, koja se zasniva na brojanom kodiranju aminokiselina i Furijeovoj transformaciji. Za predviđanje PPI proteina WT1 je korišćen integrativni algoritam, koji, pored ISM-a, obuhvata i *enrichment* analizu, mrežnu analizu i analizu koekspresije gena. Za funkcionalnu anotaciju domena proteina NISCH, identifikovanih ISM-om, korišćena je *enrichment* analiza i *GOSlims*. Razvijen je novi ISM algoritam za predikciju funkcionalnih efekata AAS, koji je evaluiran na skupu genskih varijacija koje se nalaze van konzerviranih funkcionalnih domena (engl. *not Conserved Functional Domains*, nCFD) epigenetičkih regulatora i upoređen sa najšire korišćenim programima, PolyPhen-2 i SIFT.

Filogenetski konzervirana frekvencija $F(0,00195)$ je dominantna u informacionom spektru sve četiri glavne izoforme proteina WT1. Odnos signal-šum u kros-spektru na ovoj frekvenciji veći od 10 je karakterističan za realne interakcije WT1. Skriningom UniProt baze, identifikovano je 2576 proteina sa ovom karakteristikom. Ovaj skup je statistički obogaćen proteinima uključenim u regulaciju transkripcije i obradu iRNK. U njemu se nalazi 504 proteina anotiranih terminom: 'DNK-zavisna regulacija transkripcije'. Presek ovog skupa proteina i skupa od 500 proteina dobijenog mrežnom analizom čini 51 protein.

Od njih, 36 su koeksprimirani sa WT1 u K562 ćelijskoj liniji i predstavljaju kandidate za njegove nove interaktore. Ovim analizama je humani proteom sveden na odgovarajući broj hipoteza za eksperimentalno testiranje.

Bioinformatičkom analizom su identifikovana dva funkcionalna domena NISCH-a, koji obuhvataju rezidue 680-739 i 966-1020. Računarska funkcionalna karakterizacija je pokazala da je NISCH, preko prvog domena, asociran sa: i) procesima posttranslacionih modifikacija proteina, specifično ubikvitinacijom, i ii) procesima sinaptogeneze. Preko drugog domena, NISCH je uključen u: i) transkripciju, ii) apoptozu i iii) signalnu transdukciju.

Tačnost predviđanja funkcionalnih efekata nCFD AAS u epigenetičkim regulatorima ASXL1, DNMT3A, EZH2 i TET2 PolyPhen-2 je 52%, a SIFT 57%. Zapažena niska efikasnost je najverovatnije posledica visokih korelacija predikcija ovih alata i pozicije AAS u CFD/nCFD regionima (~0,50). Razvijeni ISM algoritam je na istim podacima imao tačnost od 70%. Poređenje ova tri alata je pokazalo da je jedino ISM algoritam statistički značajno diferencirao nCFD mutacije i SNP-ove u epigenetičkim regulatorima.

Ovo istraživanje je pokazalo da se primenom računarskih alata zasnovanih na ISM-u u analizi proteina koji su uključeni u nastanak i razvoj mijeloidnih maligniteta efikasno mogu ekstrahovati informacije koje su relevantne za mehanizme i terapiju ovih bolesti.

Ključne reči: mijeloidni maligniteti, personalizovana medicina, protein-protein interakcije, proteinski domeni, aminokiselinske substitucije, metod informacionog spektra, WT1, NISCH, epigenetički regulatori.

Naučna oblast: Molekularna biologija

Uža naučna oblast: Bioinformatika

UDK broj: 616.155.392:[577.112:57.087.1](043.3)

Bioinformatics analysis of proteins involved in pathogenesis of myeloid malignancies

Abstract

Myeloid malignancies are clonal diseases of myeloid cell line, which arise, primary, as a consequence of aberrant self-renewal and differentiation of hematopoietic stem cells. Annual incidence in Europe is 7.5-8.6/100 000 and the mean 5-year survival rate, with standard therapy, is 37%. Since malignancies often carry more than one driver mutation, with high intra- and inter-cancer genetic heterogeneity, there is a need for individualized therapeutic approach to each patient. Personalized medicine is based on the principle of achieving maximal therapeutic efficacy through the use of targeted drugs for biologically characterized patients. Bioinformatics has an important role in identification of molecules for targeted therapy, rational drug design and identification of informative biomarkers for companion diagnostics.

The subject of this research included genes and their protein products that have important roles in the pathogenesis of myeloid malignancies. WT1 is a transcription factor that controls the expression of genes involved in apoptosis, proliferation and differentiation. This gene is overexpressed in 70-90%, mutated in ~10% of acute myeloid leukemia (AML) and it is engaged in more than 40 time- and context-dependant protein-protein interactions. Characterizing entire network of WT1 cofactors can improve identification of new molecules and their interactions that could be targets for therapy of AML.

Recently identified tumor suppressor in breast cancer, NISCH, affects RAC signaling pathway, which is important in regulation of hematopoiesis and has a role in pathogenesis of myeloid malignancies, so it is considered an attractive target for targeted therapy.

Characterization of functional domains of NISCH can enable design and optimization of compounds with anticancer effect.

In the last decade, the importance of epigenetic regulators is determined, which are, besides transcription factors and signaling molecules, necessary for normal functioning of hematopoietic system. These molecules have tumor suppressing effect and mutations in ASXL1, DNMT3A, EZH2 and TET2 are present in all types of myeloid malignancies and probably lead to the genesis of these diseases. Detection of these deleterious mutations is important companion diagnostic for epigenome-affecting therapeutics.

In this research, bioinformatics programs based on the Informational Spectrum Method (ISM) were applied for: i) the prediction of protein-protein interactions (PPI) of WT1, ii) the identification and characterization of NISCH functional domains and iii) the prediction of functional effects of amino acid substitutions (AAS) in epigenetic regulators ASXL1, DNMT3A, EZH2 and TET2. ISM is a method for sequence analysis, based on the numerical coding of amino acids and Fourier transform. For the prediction of WT1 PPIs, we used integrative algorithm, which contains, besides ISM analysis, enrichment, protein network and gene co-expression analyses. For the functional annotation of NISCH protein domains, identified by ISM, enrichment and GOSlims analyses were used. Finally, new ISM algorithm for prediction of functional effects of AAS was evaluated on the set of genetic variations positioned outside conserved functional domains (nCFD – not Conserved Functional Domains) of epigenetic regulators and compared with the most commonly used programs, PolyPhen-2 and SIFT.

Phylogenetically conserved frequency $F(0.00195)$ is dominant in the informational spectra of all four major WT1 protein isoforms. Signal-to-noise ratio on this frequency in cross-spectrum higher than 10 is characteristic for the real WT1 interactions. Screening of UniProt database identified 2 576 proteins with this characteristic. This set is enriched in proteins involved in transcription regulation and mRNA processing. In this dataset, 504 proteins are annotated with the term: 'DNA-templated regulation of transcription'. Overlap of this set of proteins and the set resulting from the protein network analysis contains 51

proteins. Of these, 36 are co-expressed with WT1 in the K562 cell line and represent candidates for its interacting partners. With these analyses we constricted human proteome to experimentally testable number of hypothesis.

Bioinformatics analysis identified two functional NISCH domains encompassing residues 680-739 and 966-1020. Computational functional analyses showed that NISCH associates, through the first domain, with: i) protein posttranslational modification processes, specifically ubiquitination, and ii) processes of sinaptogenesis. Through the second domain, NISCH is involved in: i) transcription, ii) apoptosis and iii) signal transduction.

Accuracy of prediction of the functional effects of nCFD AAS in epigenetic regulators ASXL1, DNMT3A, EZH2 and TET2 is 52% for the PolyPhen-2 and 57% for the SIFT. The observed low efficacy probably arises from high correlations of the predictions and the AAS position in the CFD/nCFD regions (~0.50). Developed ISM algorithm has, on the same data, accuracy of ~70%. Comparison of the efficacy of these three tools showed that only ISM algorithm statistically significantly differentiated between nCFD mutations and SNPs in epigenetic regulators.

This research showed that through the application of computational ISM-based tools for the analyses of proteins involved in occurrence and progression of myeloid malignancies one can efficiently extract information relevant for the mechanisms and therapy of these diseases.

Key words: myeloid malignancies, personalized medicine, protein-protein interactions, protein domains, amino acid substitutions, Informational Spectrum Method, WT1, NISCH, epigenetic regulators.

Scientific field: Molecular biology

Scientific discipline: Bioinformatics

UDC number: 616.155.392:[577.112:57.087.1](043.3)

Spisak skraćenica

2-HG – 2-hidroksiglutarat

5-hmC – 5-hidroksi-metilcitozin

5-mC – 5-metilcitozin

α -KG – α -ketoglutarat

AAS – aminokiselinska supstitucija (engl. *amino acid substitution*)

ak – aminokiselina

AML – akutna mijeloidna leukemija

AUC – površina ispod ROC krive (engl. *Area Under the ROC Curve*)

CDD – engl. *Conserved Domains Database*

CDK - ciklin-zavisne kinaze (engl. *Cycline Dependant Kinases*)

CFD – konzervirani funkcionalni domeni (engl. *Conserved Functional Domains*)

CHO – engl. *Chinese Hamster Ovary*

CIS – konsenzus informacioni spektar (engl. *Consensus Informational Spectrum*)

CML – hronična mijeloidna leukemija (engl. *Chronic Myeloid Leukemia*)

CMML – hronična mijelomonocitna leukemija (engl. *Chronic Myelomonocytic Leukemia*)

CTD – C-terminalni domen

DAVID – engl. *Database for Annotation, Visualization and Integrated Discovery*

DFT – diskretna Furijeova transformacija

DNMT – DNK metiltransferaze

dt – divlji tip

EASE – engl. *Expression Analysis Systematic Explorer*

EIIP – potencijal electron-jon interakcija (engl. *Electron-Ion Interaction Potential*)

ET – esencijalna trombocitemija

FFT – brza Furijeova transformacija (engl. *Fast Fourier Transform*)

GEO – engl. *Gene Expression Omnibus*

GO – engl. *Gene Ontology*

HSC – hematopoetske matične ćelije (engl. *Hematopoietic Stem Cells*)

I₁-IR – I₁ imidazolinski receptor

IAA-RP – engl. *imidazol-4-acetic acid ribotide*

IRAS – engl. *Imidazoline Receptor Antisera-Selected*

ISM – metod informacionog spektra (engl. *Informational Spectrum Method*)

JMML – juvenilna mijelomonocitna leukemija

MDS – mijelodisplastični sindromi

MDS/MPN – mijelodisplastične/mijeloproliferativne neoplazije

miRNK – mikro RNK

MPN – mijeloproliferativne neoplazije

MRD – minimalna rezidualna bolest (engl. *Minimal Residual Disease*)

MSA – višestruko poravnanje sekvenci (engl. *Multiple Sequence Alignment*)

MUT – mutacija

nCFD – engl. *not Conserved Functional Domains*

NuRD – engl. *Nucleosome Remodeling Deacetylase*

P-TEFb – engl. *Positive Transcription Elongation Factor b*

pAML – primarna AML

PcG – engl. *Polycomb-Group proteins*

PFPP – engl. *Protein Function Prediction*

PMF – primarna mijelofibroza

PolII – RNK polimeraza II

PolyPhen-2 – engl. *Polymorphism Phenotyping v2*

PPI – protein-protein interakcija

PR-DUB – engl. *Polycomb-Repressive Deubiquitylase*

PRC2 – engl. *Polycomb Repressive Complex 2*

pseAAC – engl. *pseudo Amino Acid Composition*

PSIC – engl. *Position-Specific Independent Counts*

PV – polycythemia vera

RARS – refraktorna anemija sa kružnim sideroblastima (engl. *Refractory Anemia with Ring Sideroblasts*)

RARS-T – RARS sa izraženom trombocitozom

ROC – engl. *Receiver Operating Characteristic*

RVLM – rostralna ventrolateralna medulla

S/MARs – engl. *Scaffold/Matrix Attachemnt Regions*

S/N – odnos signal-šum (engl. *Signal to Noise ratio*)

sAML – sekundarna AML

SIFT – engl. *Sorting Intolerant From Tolerant*

SNP – engl. *Single Nucleotide Polymorphism*

STRING – engl. *Search Tool for the Retrieval of Interacting Genes*

Sadržaj

1. Uvod	1
1.1. Mijeloidni maligniteti	3
1.1.1. Klasifikacija mijeloidnih maligniteta	4
1.1.2. Molekularna patogeneza mijeloidnih maligniteta	7
1.2. Personalizovana medicina.....	8
1.2.1. Protein-protein interakcije (PPI) kao mete za ciljanu terapiju	10
1.2.1.1. Uloga proteina WT1 i njegovih PPI u patogenezi mijeloidnih maligniteta.....	10
1.2.2. Ligand-vezivni domeni proteina u racionalnom dizajnu novih lekova	15
1.2.2.1. NISCH kao terapijski target i identifikacija njegovih ligand-vezivnih domena	16
1.2.3. Biomarkeri za prateću dijagnostiku.....	19
1.2.3.1. Mutacije u epigenetičkim regulatorima – markeri mijeloidnih maligniteta.....	20
1.3. Primena bioinformatike u analizi protein	24
1.3.1. Bioinformatička analiza proteina ISM metodom	26
1.3.2. Bioinformatičke metode za predikciju PPI.....	28
1.3.3. Bioinformatičke metode za predikciju funkcionalnih domena protein	31
1.3.4. Bioinformatičke metode za predikciju funkcionalnih efekata AAS.....	32
2. Ciljevi istraživanja.....	35
3. Materijal i metode	36
3.1. Podaci.....	36
3.1.1. Interaktori proteina WT1	36
3.1.2. AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2	36

3.2. Metoda informacionog spektra – ISM	37
3.2.1. Osnove ISM metode	37
3.2.2. Programi EIIP/ISM platform	42
3.2.2.1. ProteinSpektar	43
3.2.2.2. KrosSpektar	43
3.2.2.3. AKSkener	43
3.2.2.4. Mutacije i MutacijeDFT	44
3.2.2.5. FilterDFTFFTBaze	44
3.3. Ontologija gena i enrichment analiza GO termina	45
3.3.1. Ontologija gena	45
3.3.2. GO Slims i QuickGO	46
3.3.3. DAVID	46
3.4. STRING	48
3.5. Analiza koekspresije	49
3.6. Alati za ispitivanje funkcionalnih efekata polimorfizama	51
3.6.1. PolyPhen-2	51
3.6.2. SIFT	52
3.7. Statistička analiza	54
3.7.1. Unakrsne tabele	54
3.7.2. ROC krive	55
3.7.3. Korelaciona analiza	56
3.7.4. Mann-Whitney U Test	57
3.8. Algoritam za predikciju PPI	57
4. Rezultati	59
4.1. Predikcija novih PPI proteina WT1	59

4.1.1. Lista poznatih interaktora proteina WT1	59
4.1.2. Predviđanje PPI proteina WT1	61
4.1.2.1. Kros-spektralna analiza WT1	61
4.1.2.2. Enrichment analiza GO integrativnog algoritma primenjena na WT1	64
4.1.2.3. STRING analiza integrativnog algoritma primenjena na WT1	65
4.1.2.4. Integrisanje rezultata ISM, GO i STRING analize proteina WT1	67
4.1.2.5. Analiza koekspresije integrativnog algoritma primenjena na WT1	69
4.2. Predikcija i karakterizacija funkcionalnih domena proteina NISCH	71
4.2.1. Predikcija funkcionalnih domena proteina NISCH	71
4.2.2. Funkcionalna karakterizacija domena proteina NISCH	72
4.3. Predikcija funkcionalnih efekata AAS u nCFD regionima epigenetičkih regulatora ASXL1, EZH2, DNMT3A i TET2	77
4.3.1. Lista AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2	77
4.3.2. Efikasnost PolyPhen-2 i SIFT u predikciji funkcionalnih efekata nCFD AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2	80
4.3.3. ISM algoritam za predikciju funkcionalnih efekata AAS	83
4.3.3.1. Generisanje informacionih spektara	85
4.3.3.2. Računanje ISM skora	85
4.3.3.3. Selekcija značajne frekvence	85
4.3.3.4. Određivanje granične vrednosti amplitude za selektovanu frekvencu	86
4.3.3.5. Predikcija funkcionalnog efekta AAS	86
4.3.4. Predviđanje funkcionalnih efekata AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2 pomoću ISM algoritma	86
4.3.4.1. Predikcija funkcionalnih efekata AAS u proteinu ASXL1	86
4.3.4.2. Predikcija funkcionalnih efekata AAS u proteinu EZH2	88
4.3.4.3. Predikcija funkcionalnih efekata AAS u proteinu DNMT3A	89
4.3.4.4. Predikcija funkcionalnih efekata AAS u proteinu TET2	90

4.3.5. Poređenje efikasnosti ISM algoritma u predikciji funkcionalnih efekata nCFD AAS u epigenetičkim regulatorima sa PolyPhen-2 i SIFT	91
5. Diskusija.....	94
6. Zaključci.....	114
7. Literatura	116
Prilog 1	158
Prilog 2	166
Prilog 3	168
Prilog 4	171
Prilog 5	175

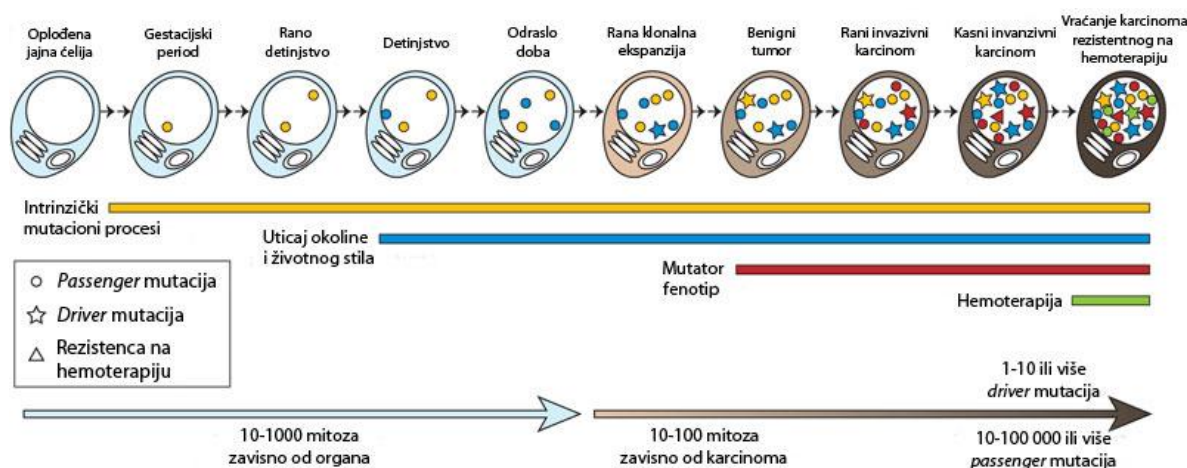
1. Uvod

Maligniteti (karcinomi, kanceri, neoplazije) su jedan od glavnih izvora mortaliteta u ljudskoj populaciji, sa 8,2 miliona smrtnih ishoda u 2012. godini (Ferlay i sar. 2013). Ove bolesti predstavljaju klonalne proliferacije, koje nastaju kao posledica mutacija koje ćelijama pružaju selektivnu prednost. Smatra se da svi maligniteti imaju istu patogenezu i osnovna svojstva, koja podrazumevaju samoodrživu proliferaciju, izbegavanje supresora rasta, suprotstavljanje ćelijskoj smrti i sticanje replikativne besmrtnosti, a neki karcinomi, pored toga, indukuju angiogenezu i dovode do nastanka metastaza (Hanahan i Weinberg 2000). Proces nastanka maligniteta je analogan Darvinovskoj evoluciji i zasniva se na: sticanju genetičkih varijacija u individualnim ćelijama kroz, manje-više, nasumične mutacije i prirodnu selekciju, koja deluje na rezultujući fenotipski diverzitet. U ljudskom organizmu se u svakom trenutku nalazi na hiljade ćelija sa stečenim varijacijama koje im omogućavaju bolje preživljavanje od susednih ćelija, ali su one, najčešće, ograničenog aberantnog rasta i nemaju klinički značaj. Međutim, povremeno, pojedinačna ćelija može da stekne skup mutacija koje joj omogućavaju autonomnu, nekontrolisanu proliferaciju i rast.

Ćelije maligniteta, kao i sve ćelije u organizmu, sadrže genetički materijal poreklom iz oplodene jajne ćelije, sa određenim brojem stečenih razlika ([Slika 1](#)). One se nazivaju somatske mutacije i razlikuju se od mutacija u polnim ćelijama – '*germline*', koje se nasleđuju od roditelja i prenose sledećoj generaciji. Somatske mutacije u genomu karcinoma se mogu podeliti na tačkaste mutacije (substitucija pojedinačne baze), insercije i delecije, hromozomske rearanžmane i aberacije broja kopija određenog dela genoma u odnosu na diploidnu strukturu.

Tačkaste varijacije su najčešće promene u ljudskom genomu (1000 Genomes Project Consortium 2012), pri čemu se u biomedicinskoj literaturi koriste različiti termini zavisno od njihovog patološkog efekta: promena koja je asocirana sa bolešću se naziva mutacija, pri čemu su u onkologiji to najčešće somatske mutacije, dok se promena koja nije asocirana sa

bolešću ili ona koja se sa određenom učestalošću uočava u opštoj populaciji naziva polimorfizam (engl. *Single Nucleotide Polymorphism*, SNP) (Human Genome Variation Society 2013). Pored toga, tačkaste varijacije u kodirajućoj sekvenci genoma mogu biti bez posledica u proteinskoj sekvenci ili mogu dovesti do aminokiselinskih supstitucija (engl. *amino acid substitution*, AAS). Osim genetičkih promena, genomi maligniteta sadrže i stečene epigenetičke promene, koje menjaju strukturu hromatina i ekspresiju gena.



Slika 1. Genetičke promene, koje dovode do patogeneze maligniteta. Prikazane su sukcesivne mitotičke deobe od oplođene jajne ćelije do maligne ćelije, sa naznačenim vremenom sticanja somatskih mutacija i procesima koji tome doprinose. Modifikovano prema Stratton i sar. 2009.

Genomi malignih ćelija sadrže dva tipa somatskih mutacija: i) vodeće – ‘*driver*’ mutacije pružaju selektivnu prednost ćelijama i uzročno su povezane sa nastankom maligniteta i ii) prateće – ‘*passenger*’ mutacije su biološki neutralne, a prisutne su u malignim ćelijama samo zato što su se nalazile u progenitorskoj ćeliji od koje je klonalnom ekspanzijom nastao karcinom. Ako se mutacije u nekom genu sreću sa značajno većom učestalošću u malignim ćelijama nego što bi se to očekivalo na osnovu pozadinske stope mutageneze, to je dobar pokazatelj kauzalne uloge tog gena u patogenezi maligniteta, pa se takvi geni nazivaju ‘kancerski geni’ (engl. ‘*cancer genes*’) (Stratton i sar. 2009). Identifikacija bioloških procesa u koje su uključeni proteinski produkti ovih ‘kancerskih

gena', funkcija koje obavljaju u ćeliji i pozicije u mreži interakcija proteina i drugih bioloških molekula, omogućava bolje razumevanje patogeneze maligniteta i razvijanje tretmana koji deluju na izmenjene molekule i procese. Takođe, 'driver' mutacije predstavljaju biomarkere za stratifikaciju maligniteta, postavljanje tačne dijagnoze i prognoze, kao i terapijsko odlučivanje. Korišćenje biomarkera i terapeutika usmerenih na aberantne molekule i procese leže u osnovi personalizovane medicine (detaljnije u Poglavlju 1.2.).

1.1. Mijeloidni maligniteti

Mijeloidni maligniteti su klonalne bolesti ćelija mijeloidne krvne loze, koja obuhvata granulocite (neutrofile, bazofile, eozinofile), monocite/makrofage, mastocite, eritrocite i megakariocite. Tu spada nekoliko grupa bolesti, sa različitim kliničkim i patološkim manifestacijama, koje su, pre svega, posledica poremećaja samo-obnavljanja i diferencijacije hematopoetskih matičnih ćelija (engl. *Hematopoietic Stem Cells*, HSC).

Mijeloidni maligniteti se karakterišu povećanim brojem progenitorskih mijeloidnih ćelija, blasta, u kostnoj srži i poremećajem njihovog sazrevanja, što rezultuje hematopoetskom insuficijencijom (Ramdass i sar. 2013). Simptomi mijeloidnih maligniteta zavise od ugrožene ćelijske linije.

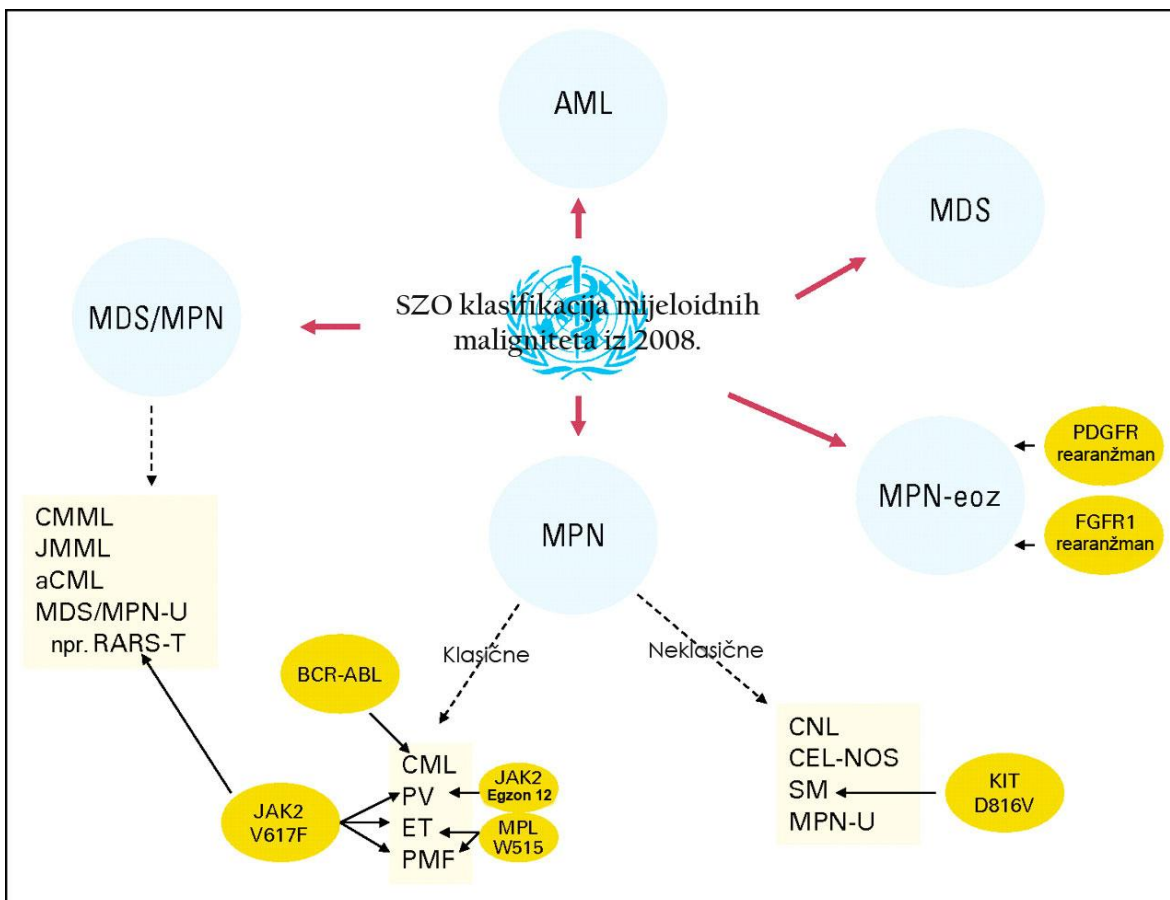
Godišnja incidenca mijeloidnih maligniteta u Evropi (standardizovana po godinama) je 7,5-8,6/100 000 (Sant i sar. 2010; Visser i sar. 2012). Za većinu ovih bolesti, incidenca je nešto viša kod muškaraca nego kod žena, a takođe raste sa godinama. Mijeloidni maligniteti su, pre svega, bolesti starenja (Appelbaum i sar. 2006), pa se smatra da će se njihova incidenca povećavati u skladu sa starenjem same ljudske populacije.

1.1.1. Klasifikacija mijeloidnih maligniteta

Klasifikacija mijeloidnih maligniteta obuhvata pet glavnih kategorija ([Slika 2](#)), koje se definišu specifičnom morfoloijom, citohemijom, imunološkim profilima, genetikom i kliničkim svojstvima (Vardiman i sar. 2009):

- i) ***Mijeloproliferativne neoplazije (MPN)*** – predstavljaju poremećaje produkcije ćelija krvi u kostnoj srži, tj. kostna srž produkuje povećan broj ćelija jedne ili više linija mijeloidne krvne loze, a bolest nastaje kao posledica nagomilavanja ovih ćelija u kostnoj srži ili krvotoku (Dameshek 1951). Najznačajniji entiteti su hronična mijeloidna leukemija (engl. *Chronic Myeloid Leukemia*, CML) i klasične MPN bolesti - polycythemia vera (PV), esencijalna trombocitemija (ET) i primarna mijelofibroza (PMF).
- ii) ***Mijeloidne i limfoidne neoplazije asociirane sa eozinofilijom i anomalijama PDGFRA, PDGFRB ili FGFR1*** – u najvećem broju slučajeva se manifestuju kao MPN, AML ili T limfoblastni limfom, uvek sa eozinofilijom (Brain 2010).
- iii) ***Mijelodisplastični sindromi (MDS)*** – karakteriše ih klonalna hematopoeza, poremećena diferencijacija (mijelodisplazija), periferna citopenija i rizik od progresije u AML (Bejar i sar. 2011 (b)). Glavni entiteti u ovoj grupi bolesti su refraktorna citopenija sa displazijom jedne linije, refraktorna citopenija sa displazijom više linija, refraktorna anemija sa kružnim sideroblastima (engl. *Refractory Anemia with Ring Sideroblasts*, RARS), refraktorna anemija sa viškom blasta i MDS sa izolovanim del(5q).
- iv) ***Mijelodisplastične/mijeloproliferativne neoplazije (MDS/MPN)*** – heterogena grupa mijeloidnih maligniteta, koje imaju neke kliničke, laboratorijske i/ili morfološke osobine specifične za MDS, a druge za MPN (Vardiman i sar. 2009). Ovi poremećaji obuhvataju hroničnu mijelomonocitnu leukemiju (engl. *Chronic Myelomonocytic Leukemia*, CMML), atipičnu CML, juvenilnu mijelomonocitnu leukemiju (JMML) i neklasifikovane MDS/MPN, od kojih je najznačajniji privremeni entitet RARS sa izraženom trombocitozom (RARS-T).

v) **Akutna mijeloidna leukemija (AML) i srodne neoplazije** – najčešći mijeloidni malignitet kod odraslih. To je heterogena klonalna bolest HSC ili blasta, koje gube sposobnost normalne diferencijacije i odgovora na normalne regulatore proliferacije, transformišući se u leukemijske matične ćelije sa povećanom sposobnošću samo-obnavljanja (Estey i Dohner 2006; Renneville i sar. 2008). Najznačajniji uticaj na tok bolesti, kao i na prognozu i odgovor na terapiju, ima citogenetički profil (Grimwade i sar. 1998; Slovak i sar. 2000; Mrózek i sar. 2001; Byrd i sar. 2002). Tako, različite hromozomske aberacije karakterišu podtipove AML (Vardiman i sar. 2009), a oko 40-50% odraslih pacijenata je sa normalnim kariotipom (Grimwade i sar. 1998; Mrózek i sar. 2001). Poseban entitet u ovoj grupi je AML asocirana sa terapijom alkilirajućim agensima, radijacijom i/ili inhibitorima topoizomerase II. U ovu grupu, takođe, spadaju mijeloidni sarkomi, mijeloidne proliferacije asocirane sa Daunovim sindromom i nespecificovane leukemije. Važno je istaći da AML može nastati *de novo* (primarna AML, pAML) ili transformacijom iz hroničnih mijeloidnih maligniteta – MPN, MDS i MDS/MPN (sekundarna AML, sAML).



Slika 2. Klasifikacija mijeloidnih maligniteta. Modifikovano prema Tefferi i sar. 2011.

AML – akutna mijeloidna leukemija; MDS – mijelodisplastični sindromi; MDS/MPN – mijelodisplastične/mijeloproliferativne neoplazije; MPN – mijeloproliferativne neoplazije; MPN-eoz – mijeloidne i limfoidne neoplazije asocirane sa eozinofilijom i anomalijama PDGFRA, PDGFRB ili FGFR1.

Prosečna petogodišnja stopa preživljavanja pacijenata sa mijeloidnim malignitetima je 37%, ali postoje značajne razlike između glavnih kategorija, pa je preživljavanje pacijenata sa MPN 63%, MDS 31%, MDS/MPN 20% i AML 17% (Maynadić i sar. 2013). Takođe, preživljavanje značajno opada sa godinama u svim glavnim kategorijama mijeloidnih maligniteta, a postoje i regionalne razlike.

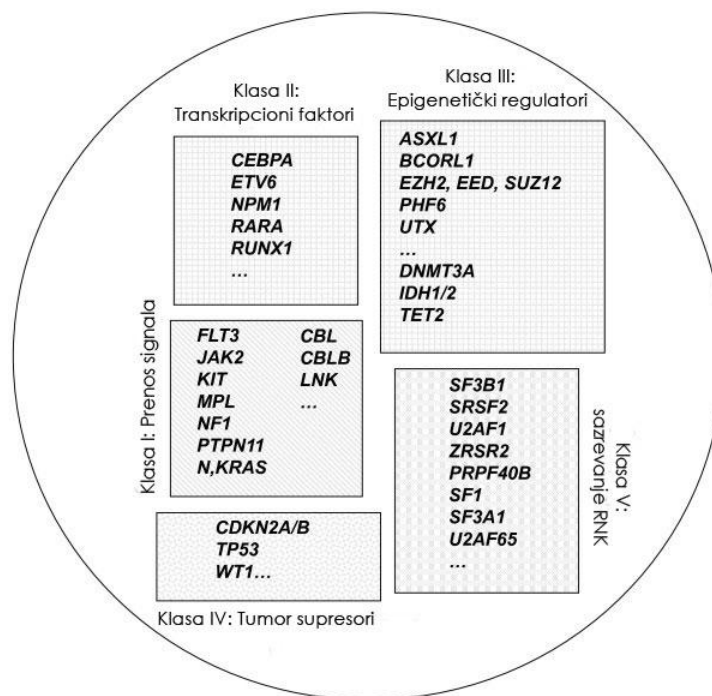
1.1.2. Molekularna patogeneza mijeloidnih maligniteta

Molekularna patogeneza je specifična za svaki od entiteta mijeloidnih maligniteta, pri čemu još uvek nisu identifikovani svi molekuli koji leže u njihovoj osnovi. Patogeneza zavisi i od toga da li entitet pripada grupi hroničnih bolesti – MPN, MDS i MDS/MPN, ili je u pitanju blastna faza/AML.

Klasičan model leukemogeneze podrazumeva minimalnu kooperaciju dve klase onkogenih događaja (engl. *two-hit model*): klasa I – mutacije u genima koji upravljaju proliferacijom (komponente signalnih puteva) i klasa II – mutacije u genima koji blokiraju diferencijaciju (uglavnom transkripcioni faktori) (Gilliland 2001). Ovaj model podrazumeva da prisustvo pojedinačne mutacije klase I može voditi nastanku MPN, dok prisustvo mutacije klase II vodi nastanku MDS (Kelly i Gilliland 2002). Međutim, sve veći broj identifikovanih mutiranih gena uključenih u patogenezu mijeloidnih maligniteta je ukazao na potrebu za revizijom *two-hit* modela (Murati i sar. 2012; Shih i sar. 2012; O'Brien i sar. 2014). Danas se smatra da je za sve mijeloidne malignitete zajedničko oštećenje dva ključna procesa – diferencijacije i samo-obnavljanja, pri čemu je najčešće oštećen i proces proliferacije (Murati i sar. 2012). Pretpostavlja se da je broj onkogenih događaja neophodnih za malignu transformaciju mnogo veći nego što se prvobitno očekivalo. Dakle, prvi korak patogeneze mijeloidnih maligniteta podrazumeva, najverovatnije, samo klonalnu ekspanziju. Geni u kojima se dešavaju mutacije odgovorne za ovaj prvi korak maligne transformacije su tumor supresori, geni uključeni u splajsovanje ili epigenetički regulatori. Sledeći koraci maligne transformacije određuju tip mijeloidnog maligniteta (Walter i sar. 2012). To podrazumeva sticanje različitog broja dodatnih mutacija usled povećane proliferacije i genetičke nestabilnosti, što vodi nastanku oligoklonalnog maligniteta.

U osnovi mijeloidnih maligniteta leže stečene genetičke anomalije i to hromozomske aberacije i/ili genske mutacije (pregled u Döhner i sar. 2010). Takođe, za ove bolesti je karakteristično prisustvo specifičnih epigenetičkih profila, profila ekspresije gena i mikro RNK (miRNK).

'Kancerski geni', čiji mutirani aleli podležu pozitivnoj selekciji, tj. predstavljaju 'driver' mutacije mijeloidnih maligniteta se mogu grupisati u pet klasa (Slika 3): i) komponente signalnih puteva, ii) transkripcioni faktori, iii) epigenetički regulatori, iv) klasični tumor supresori i v) komponente splajsozoma.



Slika 3. Shematski prikaz pet klasa gena uključenih u patogenezu mijeloidnih maligniteta. Modifikovano prema Murati i sar. 2012.

1.2. Personalizovana medicina

S obzirom da maligniteti najčešće nose više od jedne 'driver' mutacije, pri čemu je velika intra- i interkancerska genetička heterogenost (npr. Greenman i sar. 2007), postoji potreba za individualizovanim pristupom svakom pacijentu. Personalizovana medicina označava princip postizanja maksimalne terapijske efikasnosti kroz upotrebu ciljanih terapeutika kod biološki okarakterisanih pacijenata (Parkinson i sar. 2012). Cilj

personalizovane medicine je upotreba prave doze pravog leka u pravo vreme za svakog pojedinačnog pacijenta (Hamburg i Collins 2010).

Personalizaciju tretmana omogućuje i) razvoj lekova koji direktno ciljaju molekule i biološke puteve, koji uzrokuju pojavu i razvoj bolesti (ciljana terapija, engl. *targeted therapy*) i ii) razvoj dijagnostičkih metoda za prepoznavanje pacijenata, sa određenom genetičkom strukturom, koji će imati koristi od date terapije (prateća dijagnostika, engl. *companion diagnostics*). Neophodni koraci za postizanje personalizovanog pristupa su identifikacija ciljnih molekula (targeta) za razvoj terapije, sinteza malih molekula ili antitela, koja specifično deluju na target – racionalni dizajn lekova (engl. *rational drug design*), i identifikacija informativnih biomarkera, koji su relevantni za diferencijalnu dijagnozu, prognozu i terapijske odluke. Ciljana terapija za mijeloidne malignitete se intenzivno razvija u poslednje dve decenije. Terapeutici koji su odobreni za kliničku primenu, ili su u razvoju, se mogu grupisati na sledeći način:

- i) ***inhibitori signalne transdukcije***; npr. inhibitori tirozin kinaza (Druker i sar. 2001; Weisberg i sar. 2002; Talpaz i sar. 2006; Kantarjian i sar. 2007);
- ii) ***lekovi koji indukuju apoptozu***; npr. agonisti 'receptora smrti' (engl. '*death receptors*') (Ghobrial i sar. 2005);
- iii) ***regulatori ekspresije gena***; npr. inhibitori epigenetičkih regulatora DNK-metiltransferaza i histon-deacetilaza (Silverman i sar. 2002; Issa i sar. 2004; Garcia-Manero i sar. 2008).

Međutim, mijeloidni maligniteti, kao i drugi karcinomi, poseduju izuzetnu intramalignu heterogenost zbog čega se kod velikog broja pacijenata posle nekog vremena upotrebe ciljane terapije javlja rezistencija (Shah i sar. 2002; Cools i sar. 2004; Wagle i sar. 2011). Zato je neophodna identifikacija svih ključnih molekula, koji mogu biti terapijski targeti, i svih mehanizama rezistencije, a zatim koristiti specifičnu kombinovanu terapiju za svakog pojedinačnog pacijenta (Al-Lazikani i sar. 2012; Hoelder i sar. 2012; Gonzalez de Castro i sar. 2013). Takođe, neophodno je detektovati i razvijati nove prediktivne biomarkere za prateću dijagnostiku (Duffy i Crown 2013).

1.2.1. Protein-protein interakcije (PPI) kao mete za ciljanu terapiju

Svi ključni ćelijski procesi, uključujući rast, replikaciju DNK, transkripciju, translaciju i transmembranski prenos signala, regulisani su multiproteinskim kompleksima. Funkcija, aktivnost i specifičnost ovih kompleksa je kontrolisana proteinskim interakcijama među njihovim subjedinicama. Gubitak ključne interakcije ili formiranje neodgovarajućeg kompleksa je najčešći uzrok bolesti (Ryan i Matthews 2005). U karcinomima, PPI formiraju signalna čvorišta, koja prenose patofiziološke signale duž molekularnih mreža u ćelijama, i time promovišu genezu, progresiju i metastaziranje maligniteta. Zbog toga su PPI važna nova klasa molekulskih targeta za ciljanu terapiju maligniteta (Zinzalla i Thurston 2009; Ivanov AA i sar. 2013). Ovo je omogućeno velikim napretkom u poznavanju ključnih PPI uključenih u razvoj različitih tipova karcinoma i razvojem tehnologija za sintezu modulatora proteinskih interakcija (peptidomimetika i malih molekula). Najznačajnija dostignuća u ovoj oblasti su inhibitori MDM2 i TP53 interakcije (Popowicz i sar. 2011; Vu i Vassilev 2011).

1.2.1.1. Uloga proteina WT1 i njegovih PPI u patogenezi mijeloidnih maligniteta

WT1 igra centralnu ulogu u razvoju organa genito-urinarnog sistema, hematopoetskog sistema i drugih regiona embriona. Najvažnija molekulska funkcija WT1 je regulacija transkripcije, čime kontroliše ekspresiju gena uključenih u apoptozu, proliferaciju i diferencijaciju. Transkripciona funkcija WT1 je zavisna od konteksta, tj. može biti modulirana velikim brojem kofaktora (Roberts 2006). Uloga WT1 u patogenezi maligniteta je kompleksna – u nekim ima tumor supresorsku, a u nekim onkogenu ulogu (Yang L i sar. 2007).

WT1 gen je lociran na hromozomu 11p13 i sadrži dva dominantna mesta alternativnog splajsovanja, koja daju četiri glavne izoforme ovog proteina: alternativno splajsovanje egzona 5 daje izoforme sa ili bez 17 aminokiselina ($\pm 17ak$) na poziciji 250-266, a alternativno splajsovanje kraja egzona 9 daje izoforme sa ili bez 3 aminokiseline ($\pm KTS$) pozicionirane na C-kraju proteina (Call i sar. 1990; Gessler i sar. 1990). $\pm 17ak$ izoforme su

specifične za sisare, dok su \pm KTS izoforme evoluciono konzervirane kod svih ispitivanih kičmenjaka (Kent i sar. 1995). Pored alternativnog splajsovanja, alternativni start i stop kodoni, kao i RNK editovanje dovode do nastanka 36 različitih izoformi WT1.

N-terminalni domen WT1 je bogat Pro-Glu sekvencama i uključen je u interakcije WT1 sa nekim proteinima i RNK (Kennedy i sar. 1996; Nachtigal i sar. 1998; Discenza i sar. 2003; Carpenter i sar. 2004; Hartkamp i sar. 2010). Ovaj domen je ključan za regulaciju transkripcione funkcije WT1, i represije i aktivacije transkripcije. C-terminalni domen WT1 se sastoji iz četiri C₂-H₂ cink-prstića, koji omogućavaju vezivanje DNK, a uključeni su i u interakcije sa RNK i nekim proteinima (Larsson i sar. 1995; Johnstone i sar. 1996; Bardeesy i Pelletier 1998; Davies i sar. 1998; Scharnhorst i sar. 2000).

Najvažnija molekulska funkcija WT1 je u regulaciji transkripcije, pri čemu ima ulogu u aktivaciji ili represiji transkripcije (Wang ZY i sar. 1993). Samo -KTS izoforme, koje čine oko trećine WT1 proteina u ćeliji, vezuju DNK sa visokim afinitetom i funkcionišu efikasno u regulaciji transkripcije (Laity i sar. 2000). Postoji veliki broj gena čiju ekspresiju reguliše WT1, a koji su uključeni u:

- **Apoptozu** – WT1 reguliše ekspresiju nekoliko gena iz porodice BCL2 i na taj način učestvuje u regulaciji apoptoze. WT1 reguliše sam gen *BCL2* (Mayo i sar. 1999), proapoptotski gen *BAK1* (Morrison DJ i sar. 2005) ili antiapoptotski *BCL2A1* (Simpson i sar. 2006) u različitim tipovima ćelija, što ima za posledicu proapoptotsko delovanje WT1 u nekim kontekstima, tj. antiapoptotsko delovanje u drugim (Loeb 2006).
- **Diferencijaciju** – Transkripciona funkcija WT1 moduliše diferencijaciju ćelija u različitim tkivima, uključujući bubrege, srce, jetru, slezinu, itd. (Armstrong i sar. 1993). U hematopoetskom sistemu, WT1 zaustavlja ćelijski ciklus i izaziva diferencijaciju primarnih hematopoetskih progenitora, kao i humanih leukemijskih ćelijskih linija HL60 i U937 (Ellisen i sar. 2001). Jedan od glavnih medijatora ovih procesa, čiju ekspresiju reguliše WT1, je p21 – inhibitor ciklin-zavisnih kinaza (Englert i sar. 1997).
- **Proliferaciju** – Povećana ekspresija WT1 u leukemijama i drugim tipovima kancera, ukazuje na ulogu WT1 u promociji proliferacije (Miwa i sar. 1992; Inoue i sar.

1997; Oji i sar. 2002, 2003 (a), 2003 (b), 2003 (c); Ueda i sar. 2003; Oji i sar. 2004 (a), 2004 (b), 2004 (c)). Egzogeno ekspresija WT1 u CML ćelijskoj lini K562 aktivira ekspresiju pro-onkogeno MYC, što dovodi do stimulacije ćelijske proliferacije (Han i sar. 2004). Skorašnja funkcionalna analiza uticaja WT1 na proliferaciju karcinoma oralnih skvamoznih ćelija je pokazala njegovu onkogeno ulogu u ovom tipu kancera, kroz transkripcionu represiju tumor supresora CDC73 (Rather i sar. 2014).

Proteinski partneri WT1 koji zajedno učestvuju u regulaciji transkripcije se mogu podeliti u tri grupe:

- i) **Nezavisni DNK-vezivni transkripcioni faktori.** Tu spadaju opšti transkripcioni faktori, kao što su GTF2B i TBP (McKay i sar. 1999; Richard i sar. 2001). Takođe, tu su proteini iz TP53 porodice proteina: TP53 (Maheswaran i sar. 1993), TP63 i TP73 (Scharnhorst i sar. 2000), koji deluju kooperativno sa WT1 u regulaciji transkripcije. WT1 deluje kao kofaktor proteinima koji učestvuju u putu biosinteze holesterola (Rae i sar. 2004), kao i NR5A1 – faktoru transkripcije u gonadama (Nachtigal i sar. 1998).
- ii) **Kofaktori koji sami ne vezuju DNK.** U ovoj grupi su, između ostalih, CREBBP (Wang W i sar. 2001), BASP1 (Carpenter i sar. 2004), WTIP (Srichai i sar. 2004) i PAWR (Johnstone i sar. 1996).
- iii) **Epigenetički regulatori.** WT1 utiče na transkripciju i kroz interakciju sa proteinima koji učestvuju u epigenetičkim procesima, pre svega PcG grupe proteina (engl. *Polycomb-Group proteins*) (Xu B i sar. 2011). Naime, na promotoru gena PAX2, WT1 interaguje sa meninom, nakon čega regrutuje DNMT1 i PcG proteine, pri čemu fizički interaguje sa EZH2 i SUZ12, što rezultuje modifikacijama histona, koje dovode do represije transkripcije gena PAX2.

WT1 ima ulogu i u post-transkripcionom metabolizmu iRNK, koja je specifična za +KTS izoforme (Larsson i sar. 1995). WT1 direktno interaguje sa IGF2 iRNK (Caricasole i sar. 1996). Za ovu interakciju je kritičan cink-prstić 1 WT1 proteina, koji se i najviše razlikuje od cink-prstića EGR1 porodice transkripcionih faktora, kojoj WT1 pripada. Pokazano je, *in vitro*, i vezivanje WT1 i α -aktinin 1 iRNK (Morrison AA i sar. 2006).

Uloga WT1 u post-transkripcionom metabolizmu iRNK je, najverovatnije, u splajsovanju. WT1 kolokalizuje sa splajsozomima i vezuje nekoliko faktora splajsovanja (pregled u Morrison AA i sar. 2008). A Markus i saradnici su pokazali da WT1 utiče na alternativno splajsovanje iRNK kroz direktnu modulaciju efekta RBM4 faktora splajsovanja (Markus i sar. 2006).

Takođe, WT1 se kreće 'napred-nazad' (engl. *shuttle*) između jedra i citoplazme, u kojoj se nalazi 10-50% ukupne količine WT1 u ćeliji, zavisno od tipa ćelije (Niksic i sar. 2004). Deo citoplazmatičnog WT1 je asociiran sa ribonukleoproteinskim partikulama, a drugi deo je lokalizovan sa polizomima koji su u aktivnoj translaciji. +KTS WT1 izoforme mogu uticati na ekspresiju nekih gena na post-transkripcionom nivou tako što povećavaju vezivanje određenih formi iRNK za polizome (Bor i sar. 2006).

Aberacije nekih od navedenih PPI, kao i nekih još neidentifikovanih PPI proteina WT1, mogu da dovedu do poremećaja ključnih ćelijskih procesa u koje je WT1 uključen, apoptoze, diferencijacije i proliferacije. To je mehanizam koji leži u osnovi uloge WT1 u patogenezi velikog broja solidnih tumora, kao i AML (Johnstone i sar. 1996; Du i sar. 2002; Menke i sar. 2002; Cheema i sar. 2003; Reizner i sar. 2005). Dakle, targetovanje PPI proteina WT1 bi omogućilo da se isti lek koristi u različitim tipovima maligniteta. Zbog toga PPI proteina WT1, pre svega, one uključene u njegovu transkripcionu aktivnost, predstavljaju nezamenjive targete za antikancersku ciljanu terapiju.

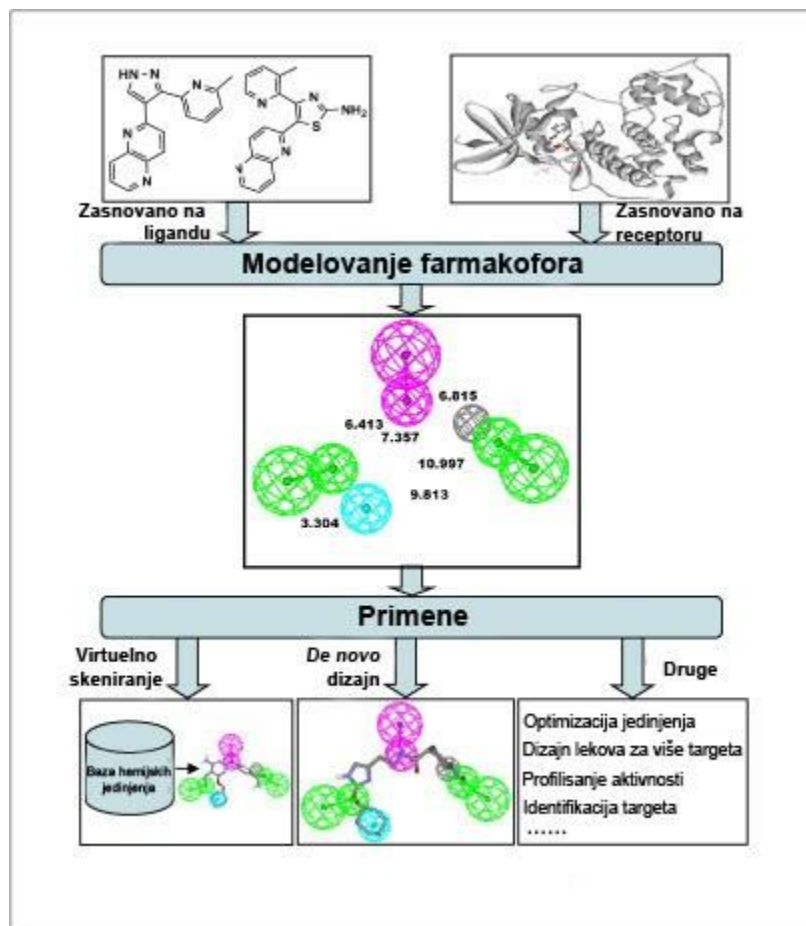
WT1 je mutiran kod ~10% pacijenata sa AML (King-Underwood i sar. 1996; King-Underwood i Pritchard-Jones 1998; Nyvold i sar. 2006; Paschka i sar. 2008; Virappane i sar. 2008; Gaidzik i sar. 2009; Hollink i sar. 2009; Renneville i sar. 2009; Ho i sar. 2010; Hou i sar. 2010). Ove mutacije su najčešće insercije ili delecije u egzonima 7 i 9, a prisutne su i mutacije bez smisla (engl. *missense*) u egzonu 9. Mutacije koje menjaju fazu okvira čitanja (engl. *frameshift*) u WT1 genu dovode do izostavljanja četiri cink-prstića na C-kraju proteina. Postoje tri moguće posledice 'skraćivanja' (engl. *truncation*) WT1: i) protein je nestabilan i biva brzo degradiran, što vodi haploinsuficijenciji; ii) protein je stabilan, ali zbog gubitka signala za nuklearnu lokalizaciju, ne obavlja svoje funkcije (engl. *loss-of-*

function), u transkripciji i obradi iRNK, u jedru; ovo, takođe, vodi haploinsuficijenciji; iii) protein je stabilan i ima novu, patološku funkciju (engl. *dominant-negative*). WT1 mutacije su najčešće u citogenetički normalnim AML ili su asocirane sa mutacijama u FLT3 (Summers i sar. 2007; Paschka i sar. 2008; Gaidzik i sar. 2009). Imaju prognostički značaj i ulogu u nastanku relapsa (Summers i sar. 2007; Paschka i sar. 2008; Gaidzik i sar. 2009; Ho i sar. 2010; Hou i sar. 2010).

U hematopoetskom sistemu, WT1 je ekspimiran u malom broju (~1%) multipotentnih progenitorskih ćelija (Hosen i sar. 2002). U normalnom razvoju hematopoetskog sistema, ekspresija WT1 je povećana tokom rane mijeloidne diferencijacije, a smanjena je u kasnijim stadijumima diferencijacije, pa je samo mali deo (<1%) diferenciranih ćelija pozitivno na WT1 (Hosen i sar. 2007). S druge strane, *WT1* gen je ekspimiran u većini (70-90%) akutnih leukemija (Miwa i sar. 1992; Menssen i sar. 1995; Inoue i sar. 1997; Østergaard i sar. 2004). Praćenje nivoa ekspresije WT1 je klinički relevantno za: i) praćenje minimalne rezidualne bolesti (engl. *Minimal Residual Disease*, MRD), što se može iskoristiti za predikciju relapsa (Inoue i sar. 1994, 1996; Cilloni i sar. 2002; Barragán i sar. 2004; Østergaard i sar. 2004; Weissner i sar. 2005; Cilloni i sar. 2009); ii) targetovanje WT1-pozitivnih leukemijskih ćelija imunoterapijom (Gao L i sar. 2000; Oka i sar. 2004; Rezvani i sar. 2008; Keilholz i sar. 2009). Ova specifična ekspresija WT1 gena u leukemijskim ćelijama predstavlja ključan argument za razmatranje PPI proteina WT1 kao novih targeta za ciljanu terapiju AML. Iako su do sada opisane neke interakcije proteina WT1, još uvek nije razjašnjena njegova 'kameleonska' uloga u normalnom razviću i patogenezi mijeloidnih maligniteta (Huff 2011). Poznavanje celokupne PPI mreže WT1 doprineće identifikaciji novih molekula i proteinskih interakcija, koje bi bili targeti za ciljanu terapiju AML.

1.2.2. Ligand-vezivni domeni proteina u racionalnom dizajnu novih lekova

U prvim fazama razvoja novih lekova vrši se virtuelni skrining biblioteka hemijskih jedinjenja, u potrazi za onima koja deluju na neki selektovani target (Jorgensen 2004). Ukoliko se vrši skrining zasnovan na strukturi receptora ([Slika 4](#)), za odabir i optimizaciju jedinjenja – kandidata za lek (engl. *lead optimization*), neophodno je poznavanje strukture ligand-vezivnih mesta na targetu (Christopoulos 2002). Kada struktura targeta nije poznata, za virtuelni skrining i optimizaciju jedinjenja se koriste modeli farmakofora (Yang SY 2010), koje predstavljaju niz svojstava neophodnih za optimalnu interakciju sa specifičnim biološkim targetom i pokretanje (ili blokiranje) njegovog biološkog odgovora (Wermuth i sar. 1998). Poznavanje funkcionalnih regiona u sekvenci targeta, na koje je potrebno terapijski delovati, može da poboljša model farmakofora za virtuelno pretraživanje biblioteka hemijskih jedinjenja. Isti principi se mogu primeniti u *de novo* dizajnu terapijskih ([Slika 4](#)).



Slika 4. Shematski prikaz primene modela farmakofora. Modifikovano prema Yang SY 2010.

1.2.2.1. NISCH kao terapijski target i identifikacija njegovih ligand-vezivnih domena

NISCH (sinonim je IRAS od engl. *Imidazoline Receptor Antisera-Selected*), strukturno jedinstven protein, sa malo sličnosti sa poznatim proteinima u bazama podataka, je jedini do sada opisani kandidat za I₁ imidazolinski receptor (I₁-IR). Pokazano je da:

- iRNK NISCH/IRAS je lokalizovana u neuronima mozga očekivano za I₁-IR (Ivanov TR i sar. 1998).

- Postoji značajna korelacija između ukupne NISCH/IRAS iRNK i mesta vezivanja I₁ imidazolina u različitim tkivima pacova (Piletz i sar. 1999).
- Transfekcija NISCH/IRAS cDNK u CHO (engl. *Chinese Hamster Ovary*) ćelije, koje ne eksprimiraju endogeno ovaj gen, dovodi do pojavljivanja visoko afinitetnih mesta vezivanja I₁ (Piletz i sar. 2000). Transfekcija humane NISCH/IRAS cDNK u PC12 ćelije pacova, koje endogeno eksprimiraju pacovski NISCH/IRAS, dovodi do povećanja mesta vezivanja I₁, a nivo ovih mesta odgovara količini proteina NISCH/IRAS (Alahari i sar. 2000; Sano i sar. 2002; Dontenwill i sar. 2003 (a)).
- Efekat na signalne procese u PC12 ćelijama, sa povećano eksprimiranim, i u CHO ćelijama, sa egzogeno eksprimiranim NISCH/IRAS, je isti kao u situaciji tretmana ćelija sa I₁ agonistima (Edwards i Ernsberger 2003; Piletz i sar. 2003; Li F i sar. 2006).
- Tretman PC12 ćelija sa NISCH/IRAS 'antisense' oligo-DNK, koji je praćen smanjenom ekspresijom ovog gena, srazmerno smanjuje ukupan broj specifičnih mesta vezivanja I₁-imidazolina na membrani, bez uticaja na afinitet vezivanja (Sun i sar. 2007). Takođe, isti tretman PC12 ćelija smanjuje fosforilaciju MAPK1/3 (ERK1/2) uzrokovanu moksonidinom, agonistom I₁-IR.
- U RVLM (rostralna ventrolateralna medula) pacova, NISCH/IRAS 'antisense' oligo-DNK značajno smanjuje ekspresiju proteina i poništava hipotenzivni efekat rilmenidina, agoniste I₁-IR, što je posledica smanjene aktivacije MAPK1/3 (ERK1/2) u RVLM pacova (Zhang J i Abdel-Rahman 2008).

Veliki broj dokaza o ulozi NISCH-a kao I₁-IR upućuju na mogućnost korišćenja I₁-IR liganada, ili nekih njima sličnih jedinjanja, u modulaciji funkcija NISCH-a. Postojeći ligandi I₁-IR, kao što su klonidin, rilmenidin i moksonidin, se koriste kao antihipertenzivi. Poznavanje funkcionalnih domena NISCH-a, koji interaguju sa ovim ligandima, bi omogućilo racionalno dizajniranje i optimizaciju lekova.

NISCH predstavlja novi 'kancerski gen', za koji je, do sada, pokazana tumor supresorska uloga u karcinomu dojke (Ding i sar. 2008, Baranwal i sar. 2011). Ekspresija NISCH-a u tumorskim tkivima dojke je niža u odnosu na normalna tkiva, pri čemu je ovaj gubitak ekspresije asociran sa gubitkom heterozigotnosti. Povećanje ekspresije NISCH-a u

ćelijama tumora dojke značajno smanjuje rast tumora i metastaze, a implantacija tumorskog ksenografa sa utišanom ekspresijom NISCH-a dovodi do značajno bržeg rasta ovog tumora u poređenju sa kontrolnim, *in vivo*. Smanjenje ekspresije NISCH-a može biti i posledica povećane metilacije promotora ovog gena (Hoque i sar. 2008; Ostrow i sar. 2010).

NISCH postiže tumor supresorski efekat kroz uticaj na Rac signalni put. RAC1 je član porodice Rho GTPaza, ključnih za (pregled u Mack i sar. 2011; Bid i sar. 2013):

- i) **Kontrolu citoskeleta** i, posledično, fagocitoze, ćelijske migracije, rasta aksona, adhezije i diferencijacije različitih tipova ćelija. RAC1 ovo postiže kroz interakciju sa PAK porodicom kinaza (PAK1-3). Aktivirane PAK, dalje, fosforilišu i aktiviraju LIM kinaze, LIMK1 i LIMK2, čiji je supstrat kofilin. Kofilin kontroliše polimerizaciju aktina.
- ii) **Regulaciju ćelijske proliferacije i apoptoze** kroz učešće u signalnoj transdukciji uključenoj u kontrolu ekspresije nekih gena. RAC1/PAK kaskada reguliše aktivnost MAPK1/3 (ERK1/2) kinaza, JNK i p38 MAP kinaza.

NISCH utiče na sve elemente Rac signalnog puta. U ćelijama karcinoma dojke, NISCH interaguje sa citoplazmatičnim repom $\alpha 5$ integrinske subjedinice fibronektinskog receptora, koji se nalazi uzvodno od RAC1 (Alahari i sar. 2000; Alahari 2003; Baranwal i sar. 2011). Ova interakcija dovodi do aktivacije RAC/PAK/LIMK/kofilin signalnog puta i dovodi do povećane migracije epitelnih ćelija i metastaziranja karcinoma. Takođe, NISCH direktno interaguje i reguliše aktivnost FAK, PAK1 i LIMK1 proteina (Ding i sar. 2008; Baranwal i sar. 2011). NISCH deluje i na MAPK signalnu kaskadu i regulaciju apoptoze, pri čemu je ovaj efekat zavisao od tipa ćelije. I_1 -IR ligandi dovode do inhibicije norepinefrin indukovane fosforilacije p38 MAPK, dok je fosforilacija AKT sačuvana u kardiomiocitima, a inhibirana u fibroblastima (Aceros i sar. 2011). Ovo ima za posledicu anti-apoptotski efekat moksonidina na kardiomiocite, odnosno, pro-apoptotski efekat na fibroblaste.

Rac signalni put ima ključnu ulogu u regulaciji normalne hematopoeze (Gu i sar. 2003). U hematopoetskim ćelijama, RAC proteini integrišu signale uključene u regulaciju ekspresije gena, proliferacije, apoptoze, adhezije i pokretljivosti ćelija (Cancelas i Williams

2009; Karlsson i sar. 2009). Ovaj signalni put je uključen u patogenezu mijeloidnih maligniteta – CML, AML i MPN:

- Rac signalni put je ključni regulator BCR-ABL zavisne leukemogeneze CML i predstavlja važan target za ciljanu terapiju (Skorski i sar. 1998; Harnois i sar. 2003; Thomas EK i sar. 2007, 2008). Efektori BCR-ABL se preklapaju sa signalnom kaskadom nizvodno od RAC1, koja učestvuje u regulaciji proliferacije i apoptoze ćelija – ERK, AKT, FAK, JNK, p38, SRC.
- JNK ima ključnu ulogu u indukovanoj apoptozi u BCR-ABL pozitivnoj CML ćelijskoj liniji K562 (Mc Gee i sar. 2002).
- Ekspresija Rac1 i Cdc42, takođe Rho GTPaze, je povećana u životinjskom modelu AML sa MLL rearanžmanom (Somerville i Cleary 2006). U AML ćelijskim linijama sa MLL rearanžmanom, specifični inhibitori Rac signalnog puta utiču na proliferaciju, preživljavanje i progresiju kroz ćelijski ciklus (Müller i sar. 2008; Hinterleitner i sar. 2013).
- PAK kinaze su efektori onkogenih mutanata KIT u MPN i AML (Martin i sar. 2013).

Dakle, za veliki broj proteina iz Rac signalnog puta koji su uključeni u patogenezu mijeloidnih maligniteta, pokazano je da ih reguliše NISCH u srčanim ćelijama, karcinomima dojke i kolona. Iz ovoga sledi pretpostavka da NISCH, preko Rac signalnog puta, ima ulogu i u mijeloidnim malignitetima i da predstavlja novi potencijalni target za terapijsko delovanje.

1.2.3. Biomarkeri za prateću dijagnostiku

Prateća dijagnostika podrazumeva upotrebu biomarkera u kombinaciji sa određenom terapijom radi predviđanja toksičnosti i/ili odgovora na terapiju kod svakog individualnog pacijenta. Posebno je značajna u onkologiji zbog toga što (Duffy i Crown 2013): i) karcinomi često brzo napreduju i potrebno je brzo donositi terapijske odluke; ii) postoje tretmani koji su efikasni samo kod određenih grupa pacijenata; iii) antikancerska terapija je često visoko toksična; iv) mnogi antikancerski tretmani su skupi.

1.2.3.1. Mutacije u epigenetičkim regulatorima – markeri mijeloidnih maligniteta

Epigenetika se odnosi na nasledne procese, koji su nezavisni od sekvence DNK, a koji regulišu gensku ekspresiju. Epigenetičke modifikacije su ključne za genetičko programiranje i ćelijsku diferencijaciju (Mikkelsen i sar. 2007). Epigenom obuhvata modifikacije DNK (metilaciju/demetilaciju) i različite hemijske modifikacije histona.

U mijeloidnim malignitetima, postoje specifični profili metilacije DNK i modifikacija histona, koji se mogu koristiti u klasifikaciji i dijagnostici mijeloidnih maligniteta, a predstavljaju i targete za ciljanu terapiju (Alvarez i sar. 2010; Figueroa i sar. 2010 (a); Griffiths i sar. 2010). Za ove profile su odgovorne somatske alteracije u epigenetičkim regulatorima, DNMT3A, TET2, IDH1/2, EZH2 i ASXL1. Detekcija ovih štetnih mutacija je ključni prateći dijagnostik za epigenetički zasnovane terapeutike (pregled u Fathi i Abdel-Wahab 2012).

Kod sisara, metilacija DNK se dešava na poziciji C5 citozina, prvenstveno u CpG dinukleotidima i posredovana je enzimima iz porodice DNK metiltransferaza (Bestor 2000). Pored metilacije, relativno skoro je otkriven i drugi oblik epigenetičke modifikacije DNK, koji je posledica oksidacije metil grupe 5-metilcitozina (5-mC). Ovaj proces je posredovan proteinima iz TET porodice i rezultuje stvaranjem 5-hidroksi-metilcitozina (5-hmC), što najverovatnije vodi demetilaciji DNK (Tahiliani i sar. 2009). Globalna DNK hipometilacija u kancerogenezi je povezana sa genomskog nestabilnošću, povećanom stopom spontanih mutacija i potencijalnom ekspresijom transpozona (Ehrlich 2009). Hipermetilacija CpG ostrvaca u pojedinačnim promotorima vodi utišavanju gena, kao što su tumor supresori p15, p16, E-kadherini, SOCS-1, itd. (Graff i sar. 1995; Sutherland i sar. 2004; Villa i sar. 2004; Baylin 2005; Galm i sar. 2006; Boulton i Wainscoat 2007). Genski lokusi hipermetilovani u karcinomima, pripadaju potpisu matičnih ćelija (engl. *stem cell signature*) i održavaju stanje sa nekontrolisanim rastom i samo-obnavljanjem (Widschwendter i sar. 2006).

Kod sisara, postoje tri aktivne DNK metiltransferaze (DNMT) – DNMT1, DNMT3A i DNMT3B, kao i katalitički neaktivna DNMT3L. DNMT1 ima ulogu u održavanju

metilacije DNK tokom deobe, dok DNMT3A i DNMT3B imaju ključnu ulogu u *de novo* metilaciji DNK (Okano i sar. 1999). DNMT enzimi prenose metil grupu na nemodifikovani citozin koji je pozicioniran u CpG ostrvcima, što rezultuje 5-mC. Osnovna uloga metilacije DNK je u utišavanju gena (engl. *gene silencing*) (Klose i Bird 2006).

U **DNMT3A** genu su, u mijeloidnim malignitetima, detektovane mutacije bez smisla (engl. *nonsense*), *frameshift* i *missense* mutacije, a od poslednjih su najčešće mutacije na poziciji R882. Pored njih, uočen je veliki broj *missense* mutacija, od kojih su neke rekurentne. DNMT3A mutacije koje su prisutne kod 12-35% AML pacijenata (pregled literature u Prilogu 5) asocirane su sa normalnim kariotipom, kao i mutacijama u NPM1, IDH1/2 i FLT3-ITD (Ley i sar. 2010; Stegelmann i sar. 2011; Thol i sar. 2011 (a); Yan XJ i sar. 2011; Fernandez-Mercado i sar. 2012; Fried i sar. 2012; Hou i sar. 2012; Marková i sar. 2012; Renneville i sar. 2012). Ovaj gen je, takođe, mutiran i u manjem broju pacijenata sa MPN, MDS i CMML (v. Prilog 5), kao i kod nekih pacijenata sa limfoidnim malignitetima (Couronné i sar. 2012; Grossmann i sar. 2013). DNMT3A mutacije imaju važan prognostički značaj u AML i MDS, pri čemu su asocirane sa lošijom prognozom; tj. nižom stopom preživljavanja i većom stopom relapsa (Ley i sar. 2010; Lin i sar. 2011; Shen Y i sar. 2011; Thol i sar. 2011 (a), 2011 (c); Walter i sar. 2011; Yan XJ i sar. 2011; Hájková i sar. 2012; Hou i sar. 2012; Marková i sar. 2012; Renneville i sar. 2012). Mutacije u DNMT3A mogu da vode i terapijske odluke s obzirom da predviđaju odgovor na idarubicin (LaRoche i sar. 2011), daunorubicin (Patel i sar. 2012) i kombinovanu terapiju lenalidomidom i azacitidinom (Sekeris i sar. 2012). S obzirom da pokazuju stabilnost u toku evolucije AML, DNMT3A mutacije su i potencijalni biomarkeri MRD kod AML pacijenata (Hou i sar. 2012).

TET2 je član TET porodice gena (TET1-3), čiji produkti funkcionišu kao Fe(II)- i α -ketoglutarat (α -KG)-zavisni enzimi, koji katalizuju konverziju 5-mC u 5-hmC u DNK (Tahiliani i sar. 2009; Ko i sar. 2010). 5-hmC blokira vezivanje proteina koji vezuju metilovanu DNK, a koji su normalno izvor transkripcionog utišavanje genske ekspresije i dovode do povećane genske ekspresije (Pastor i sar. 2011). Pored toga, 5-hmC može voditi pasivnoj demetilaciji DNK (Valinluck i Sowers 2007). DNMT1 ne prepoznaje ovako

modifikovanu DNK, pa se metilacija gubi tokom sledećih ciklusa replikacije jer ne dolazi do metilacije citozina u novonastalim lancima DNK. Takođe, 5-hmC vodi i aktivnoj demetilaciji DNK, koju obavlja APOBEC put za reparaciju DNK, pri čemu se 5-hmC konvertuje u 5-hmU, što aktivira baznu ekscizionu reparaciju DNK u kome različiti enzimi generišu nemetilovan citozin (Guo JU i sar. 2011). TET enzimi, takođe, prevode 5-mC u 5-formilcitozin i 5-karboksicitozin, ali je biološka funkcija ovih modifikacija još uvek nejasna (Ito S i sar. 2011).

TET2 je mutiran u svim tipovima mijeloidnih maligniteta, sa učestalošću 5-13% u klasičnim MPN, oko 11% u CML, 12-26% u MDS, 42-51% u CMML i oko 8-27% u AML (pregled literature u Prilogu 5). Do sada je pokazan klinički značaj TET2 mutacija kao prognostičkih markera u citogenetički normalnoj AML i CMML, ali ne i u MDS i MPN (Bejar i sar. 2011 (a); Grossmann i sar. 2011; Metzeler i sar. 2011 (a); Patel i sar. 2012). TET2 mutirane ćelije mogu biti u osnovi relapsa kod AML pacijenata, pa predstavljaju potencijalni marker za MRD (Jan i sar. 2012). Prisustvo TET2 mutacija asocirano je sa stopom odgovora na azacitidin (Itzykson i sar. 2011).

Enzimi **IDH1** i **IDH2** katalizuju ključni korak u Krebsovom ciklusu u kome se izocitrat prevodi u α -ketoglutarat. IDH1 funkcioniše u citoplazmi i peroksizomima, a IDH2 u mitohondrijama. Mutacije u IDH1 i IDH2 su heterozigotne i uvek se dešavaju na jednom od tri visoko konzervirana mesta u njihovim sekvencama: IDH1 – R132, IDH2 – R172 i R140 (Balss i sar. 2008; Mardis i sar. 2009; Yan H i sar. 2009; Abbas i sar. 2010; Paschka i sar. 2010; Tefferi i sar. 2010). R132 mutacija u IDH1 smanjuje enzimatsku aktivnost ovog proteina (Yan H i sar., 2009; Zhao i sar., 2009), a mutirani protein stiče i novu funkciju (engl. *gain-of-function*), u aktivnoj produkciji metabolita 2-hidroksiglutarata (2-HG) (Dang i sar. 2009). 2-HG narušava funkciju α -KG-zavisnih enzima, kao što je TET2 (Figueroa i sar. 2010 (b)). IDH1/2 i TET2 mutacije se međusobno isključuju kod pacijenata sa mijeloidnim malignitetima.

Modifikacije histona imaju ključnu ulogu u svim biološkim procesima koji uključuju strukturu DNK i ekspresiju gena. Tu spadaju transkripcija, replikacija, rekombinacija DNK

i popravka oštećenja. Modifikacije histona ispoljavaju svoje efekte kroz dva glavna mehanizma (pregled u Bannister i Kouzarides 2011): i) direktan uticaj na strukturu hromatina i ii) regulaciju vezivanja efektorskih molekula. Glavni tipovi modifikacije histona su acetilacija, fosforilacija i metilacija.

EZH2 je H3K27 metiltransferaza i pripada proteinima PcG grupe – član je PRC2 kompleksa (engl. *Polycomb Repressive Complex 2*) (Schuettengruber i sar. 2007). Ovo su transkripcioni represori, ključni za regulaciju ćelijske diferencijacije i održavanje ćelijskog identiteta u različitim kontekstima (Ezhkova i sar. 2009). EZH2 ima važne uloge u inicijaciji i održavanju transkripcionog utišavanja, kroz specifičnu metilaciju histona H3 (Simon i Kingston 2009).

Aberacije EZH2 uključene u patogenezu maligniteta obuhvataju njegovu poremećenu ekspresiju i somatske heterozigotne mutacije (Varambally i sar. 2002; Bracken i sar. 2003; Varambally i sar. 2008; Ernst i sar. 2010; Nikoloski i sar. 2010). U limfomima su identifikovane aktivirajuće EZH2 mutacije na poziciji Y641 (Morin i sar. 2010), za koje je pokazano da vode povećanoj di- i trimetilaciji H3K27, uprkos poremećenoj monometilaciji (Sneeringer i sar. 2010). U mijeloidnim malignitetima je detektovan širok spektar *missense*, *nonsense* i *frameshift* mutacija u ovom genu. One se pojavljuju u 3-6% MPN, 5-6% MDS i 4-12% MDS/MPN, a retke su u AML (pregled literature u Prilogu 5). EZH2 mutacije imaju prognostički značaj u CMML-u (Grossmann i sar. 2011).

ASXL1, zajedno sa BRCA1 proteinom, pripada kompleksu PR-DUB (engl. *Polycomb-Repressive Deubiquitylase*), koji uklanja ubikvitin sa H2AK119 (Scheuermann i sar. 2010). ASXL1 učestvuje dvojako u regulaciji transkripcije – asociiran je i sa represivnim PcG kompleksom i sa aktivirajućim Trithorax kompleksom (Baskind i sar. 2009). Smatra se da ovaj protein ne poseduje enzimatsku aktivnost (Aravind i Iyer 2012). Gubitak ASXL1 u mijeloidnim hematopoetskim ćelijama, kroz mutacije koje dovode do gubitka funkcije (engl. *loss-of-function*), vodi izostajanju markera transkripcione represije – trimetilacije H3K27 (Abdel-Wahab i sar. 2012). Abdel-Wahab i sar. su pokazali da gubitak ASXL1 zaista vodi povećanju ekspresije nekih gena, uključujući HOXA gene. Izgleda da je ASXL1

važan za regrutovanje i/ili stabilnost PRC2 kompleksa u mijeloidnim hematopoetskim ćelijama, pa njegov gubitak vodi smanjenom vezivanju PRC2 kompleksa za target gene i posledično do povećanja njihove ekspresije.

Loss-of-function mutacije u ASXL1 su detektovane u skoro svim tipovima mijeloidnih maligniteta, u 5-15% klasičnih MPN, oko 15% CML, 14-22% MDS, 34-47% CMML i 3-17% AML (pregled literature u Prilogu 5). Ove mutacije pokazuju prognostički značaj u MDS, CMML i AML (Gelsi-Boyer i sar. 2010; Bejar i sar. 2011 (a); Metzeler i sar. 2011 (b); Thol i sar. 2011 (b); Gelsi-Boyer i sar. 2012; Pratcorona i sar. 2012; Schnittger i sar. 2013; Patnaik i sar. 2014).

1.3. Primena bioinformatike u analizi proteina

Bioinformatika je naučna disciplina koja istražuje biološke informacije uz primenu kompjuterskih tehnika (Luscombe i sar. 2001). Ova molekularno biološka disciplina obuhvata široki spektar istraživanja, od strukturne biologije, genomike do ispitivanja genske ekspresije. Tri glavna cilja bioinformatike su: i) jednostavna organizacija podataka, koja omogućava istraživačima pristup postojećim informacijama i objavljivanje novih podataka, ii) razvoj alata i sredstava za analizu ovih podataka, što zahteva razumevanje i kompjuterske teorije i biologije i iii) efikasna primena, koja omogućava analizu bioloških podataka i interpretaciju ovako dobijenih rezultata na biološki značajan način. Podaci koji se najčešće analiziraju bioinformatičkim metodama su DNK i proteinske sekvence, strukture makromolekula, genske ekspresije, čitavi genomi i proteomi, a primenjuju se i *textmining* tehnike.

Najviše korišćena analiza primarnih struktura proteina obuhvata pretraživanje sekvenci po sličnosti za identifikaciju homologih sekvenci i njihovo poravnanje – MSA (engl. *Multiple Sequence Alignment*). Za detekciju homologa se koriste programi BLAST i PSI-BLAST (Altschul i sar. 1997), FASTA (Pearson i Lipman 1988), HMMER3 (Johnson i sar. 2010) i drugi. Koncept homologije podrazumeva da sekvence, koje su sličnije nego što bi

se očekivalo po principu slučajnosti, dele istog evolucionog pretka. Homologe sekvence koje predstavljaju 'isti gen' u različitim vrstama se nazivaju ortolozi, a homologe sekvence nastale duplikacijom gena su paralozi. Funkcionalna karakterizacija proteina na osnovu primarne strukture se zasniva na pretraživanju sekvenci po sličnosti i identifikaciji ortologih sekvenci (Koonin 2005). MSA poravnanje homologih sekvenci omogućava detekciju visoko konzerviranih regiona proteina – funkcionalnih domena (engl. *Conserved Functional Domains*, CFD). CFD čine aktivne domene enzima, mesta direktnog vezivanja kofaktora, ligand-vezivne domene receptora i imaju druge ključne uloge u funkcionisanju proteina. S obzirom da mutacije u ovim regionima proteina najčešće narušavaju njegovu funkcionalnost, MSA-zasnovane metode se koriste i za predikciju funkcionalnih efekata AAS (Ng i Henikoff 2003; Bao i sar. 2005; Ferrer-Costa i sar. 2005; Stone i Sidow 2005; Capriotti i sar. 2006; Bromberg i Rost 2007; Adzhubei i sar. 2010).

Međutim, postoje brojni dokazi da kod mnogih proteina postoje konzervirane, funkcionalno značajne, periodičnosti koje ne odgovaraju konkretnoj aminokiselinskoj sekvenci, već su asocirane sa određenim fizičko-hemijskim karakteristikama aminokiselina (v. Poglavlje 1.3.3.). Ove 'neočigledne' periodičnosti u proteinskoj sekvenci se mogu detektovati matematičkim modelovanjem (Ifeachor i Jervis 2002). Ono podrazumeva kodiranje aminokiselina u sekvenci brojevima koji predstavljaju vrednosti nekog fizičko-hemijskog parametra i konverzijom ovog niza brojeva korišćenjem: i) *wavelet* ili ii) Furijeove transformacije. Različita fizičko-hemijska i biohemijska svojstva aminokiselina (Kawashima i sar. 2008) se mogu grupisati u pet klastera visoko međusobno koreliranih fizičko-hemijskih parametara (Atchley i sar. 2005): i) faktor I je predstavljen svojstvima kao što su polarnost i hidrofobnost aminokiselina, ii) faktor II obuhvata svojstva koja su povezana sa sekundarnom strukturom, iii) faktor III se odnosi na veličinu i zapreminu aminokiselina, iv) faktor IV reflektuje relativnu aminokiselinsku kompoziciju u različitim proteinima i v) faktor V se odnosi na elektrostatička i elektro-dinamička svojstva aminokiselina. Najčešće korišćene metode za MSA-nezavisnu analizu proteinskih sekvenci su zasnovane na pseAAC (engl. *pseudo Amino Acid Composition*) (Chou 2011), i ISM (engl. *Informational Spectrum Method*) (Veljkovic V i sar. 2007).

1.3.1. Bioinformatička analiza proteina ISM metodom

Metoda informacionog spektra, ISM, je metoda za analizu sekvenci, koja se primenjuje u istraživanju odnosa strukture i funkcije proteina i nukleotidnih sekvenci, PPI i protein-DNK interakcija (Veljković V i sar. 1985). Fizička osnova ISM se zasniva na konceptu Herberta Freliha o dalekosežnim međumolekulskim interakcijama (Fröhlich 1968, 1970, 1975). Ovaj model podrazumeva interakcije molekula u biološkim sistemima, koje deluju na rastojanjima do 1000 Å i koje se zasnivaju na rezonantnoj komunikaciji između interagujućih molekula.

Ćelija predstavlja izuzetno kompleksan sistem, koji se sastoji od stotina hiljada bioloških makromolekula, čije konstantne međusobne interakcije održavaju funkcionisanje ćelije. Interakcije su izuzetno specifične, pa npr. neki protein prepoznaje jedan ili mali broj svojih partnera među stotinama hiljada kandidata (Carbonell i sar. 2009). U isto vreme, za mnoge proteine, proces prepoznavanja je brz, što ukazuje na prisustvo dalekosežnih sila, pomoću kojih se interagujući partneri prepoznaju i međusobno privlače. Naime, interakcija dva proteina je ograničena korakom kojim se oni pronalaze u ćeliji pomoću difuzije. Pored toga, da bi formirali kompleks, proteini u trenutku susreta moraju da imaju odgovarajuću orijentaciju. Teorijski modeli, koji uzimaju ova dva faktora u obzir, predviđaju da je stopa PPI 10^5 - $10^6 \text{M}^{-1} \text{s}^{-1}$ (Northrup i Erickson 1992; Zhou 1997; Schlosshauer i Baker 2004), što je i eksperimentalno pokazano za neke interakcije (npr. Noble i sar. 1969; Foote i Eisen 1995; Hoffman i sar. 1999). Međutim, u mnogim biološkim procesima odigravaju se značajno brže asocijacije proteina, posredovane dalekosežnim prepoznavanjem, što je i eksperimentalno zapaženo za npr. dimerizaciju insulina (Koren i Hammes 1976), interakciju citohroma c i citohrom c peroksidaze (Kang i sar. 1978), peptidnih liganada i acetilholinesteraze (Radić i sar. 1997), Cdc42 i Was proteina (Hemsath i sar. 2005).

Parametar koji opisuje dalekosežne osobine bioloških molekula je potencijal elektron-jon interakcija (engl. *Electron-Ion Interaction Potential*, EIIP) (Veljkovic V 1980). EIIP predstavlja meru raspodele energije valentnih elektrona u molekulu, koja leži u osnovi dipolnog i kvadrapolnog momenta, jonizacionog potencijala, elektrofilnosti molekula, itd.

ISM procedura se sastoji iz dva koraka: i) transformacija aminokiselinske sekvence u sekvencu brojeva, pri čemu se svakoj aminokiselini dodeljuje odgovarajuća vrednost EIIP; ii) konverzija ovako dobijenog niza brojeva, pomoću Furijeove transformacije, u informacioni spektar. Furijeova transformacija je matematički alat za identifikaciju periodičnih strukturnih obrazaca u proteinskim sekvencama. Pokazano je da frekvence u informacionom spektru proteina koreliraju sa njegovim biološkim funkcijama. Dakle, ISM se koristi za detekciju latentnih periodičnosti, tj. neočiglednih strukturnih obrazaca u proteinskim sekvencama, a kompjuterski alati zasnovani na ISM algoritmu se primenjuju za:

- Analizu odnosa strukture i funkcije proteina i identifikaciju funkcionalno značajnih domena, koji predstavljaju kandidate za mesta terapijskog delovanja (Cosic i sar. 1989; Cosic i Hearn 1991; de Trad i sar. 2000; Murakami 2000; Ivanović i sar. 2004; Sauren i sar. 2004). ISM je primenjen u identifikaciji strukturnih domena proteina hemaglutinina virusa svinjskog gripa H1N1 i ptičijeg gripa H5N1 i u detekciji AAS u ovim proteinima koje modulišu interakciju hemaglutinina sa ljudskim receptorom (Veljkovic V i sar. 2009 (a), 2009 (b)).
- Kreiranje *in silico* interaktoma i pretraživanje mogućih interaktora određenih proteina i nukleotidnih sekvenci u bazama sekvenci. ISM je korišćen za predviđanje interakcija proteina i DNK i PPI (npr. Ćosić i sar. 1986; Veljković V i Metlas 1988; Skerl i Pavlović 1988; Huang J i sar. 2005; Doliana i sar. 2008; Tintori i sar. 2010). Predviđanje PPI uključenih u CML je doprinelo boljem razumevanju molekularnog mehanizma patogeneze ove bolesti (Mancini i sar. 2009, 2012).
- Procenu bioloških efekata AAS. Narušavanje periodičnosti u proteinskoj sekvenci može imati posledice na funkciju i biološku aktivnost. Pokazano je da je ISM odgovarajući metod za mapiranje ovih poremećaja i uspešno je primenjen za definisanje kriterijuma za detekciju štetnih mutacija u proteinima LPL i CX3CR1, uključenim u nastanak kardiovaskularnih bolesti i multiple skleroze, redom (Glisic i sar. 2008; Stojković i sar. 2012). TP53 je važan tumor supresor, čija aktivnost je regulisana njegovom interakcijom sa MDM2 proteinom. Definisan je ISM kriterijum za procenu efekata TP53 mutacija na ovu

interakciju (Veljkovic N i Perovic 2006). ISM je iskorišćen i za razvoj novog algoritma za praćenje efekata mutacija u hemaglutininu H5N1 virusa (Perovic i sar. 2013).

➤ De novo dizajn biološki aktivnih peptida (Krsmanovic i sar. 1998; Veljkovic N i sar. 2003).

1.3.2. Bioinformatičke metode za predviđanje PPI

Većina bioloških procesa u živim organizmima uključuje veliki broj različitih proteina, koji najčešće funkcionišu kao komponente stabilnih ili dinamičkih proteinskih kompleksa. Interaktom predstavlja skup svih molekularskih interakcija u određenoj ćeliji. Svaki višćelijski organizam sadrži više različitih interaktoma, koji su zavisni od tkiva, biološkog stanja, itd. (Bossi i Lehner 2009; Ellis i sar. 2012). Dakle, informacije o PPI su izuzetno važne za razumevanje komplikovanih molekularskih mehanizama funkcionisanja živih sistema. Takođe, narušavanje PPI može voditi razvoju bolesti, pa je identifikacija proteina čije su interakcije poremećene u nekoj bolesti važno jer oni (molekuli i njihove PPI) predstavljaju atraktivne mete za razvoj novih terapija. Rezultati interaktomskih studija su doprineli razvoju inovativnog pristupa dizajnu lekova (peptidomimetika i malih molekula), po kojem je cilj leka da blokira formiranje aberantnog proteinskog kompleksa, a ne samo uticaj na individualne ciljne proteine (Zinzalla i Thurston 2009; Ivanov AS i sar. 2011; Ivanov AA i sar. 2013).

Proteinska interaktomika je nastala na preseku različitih naučnih disciplina: biohemija, genomika, bioinformatika, kompjutersko molekularno modelovanje, ćelijska i molekularna biologija i biofizika. U interaktomici postoji veliki broj metodoloških pristupa, koji se mogu podeliti na bioinformatičke i eksperimentalne (pregled u Ivanov AS i sar. 2011). Eksperimentalna grupa metoda obuhvata metode zasnovane na genomskom pristupu, kao što su npr. dvo-hibridni sistem kvasca (engl. *yeast two-hybrid system*) i korelativni profili iRNK ekspresije, i metode zasnovane na biohemijskom pristupu, npr. metode zasnovane na afinitetnoj hromatografiji i identifikaciji proteina masenom spektrometrijom. Bioinformatička grupa metoda obuhvata pristupe zasnovane na: i) genomskim

informacijama, ii) evolucionim odnosima proteina, iii) trodimenzionalnim strukturama proteina, iv) informacijama o proteinskim domenima i v) primarnim strukturama proteina. Iako razvijene eksperimentalne tehnike mogu da se koriste na proteomskom nivou i da produkuju velike količine podataka, one i dalje pokazuju nisku pokrivenost interakcija i sklonost ka određenim tipovima proteina i ćelijskoj lokalizaciji, zbog čega je neophodan razvoj kompjuterskih metoda za predikciju PPI (Shoemaker i Panchenko 2007). Bioinformatičke metode se primenjuju u: i) odabiru potencijalnih proteinskih meta za eksperimentalni skrining, ii) validaciji eksperimentalnih podataka i iii) pružanju dodatnih informacija o interakciji, koje nisu očigledne iz eksperimentalnih tehnika, kao što je npr. predviđanje interagujućih domena.

Veze između proteina se često predstavljaju u vidu mreža, koje omogućavaju jedinstveni pogled na genom kao kompleksnu i dinamičku organizaciju, a ne samo statični skup genetičkih funkcija. Eksperimentalni podaci o direktnim, fizičkim PPI su uskladišteni u bazama podataka, kao što su DIP (Xenarios i sar. 2002), BioGRID (Stark i sar. 2006), MatrixDB (Chautard i sar. 2009), IntAct (Hermjakob i sar. 2004) i MINT (Zanzoni i sar. 2002). Informacije o ćelijskim putevima u koje su uključene PPI su uskladištene u KEGG (Kanehisa i Goto 2000) i Reactome (Joshi-Tope i sar. 2005) bazama. STRING je servis koji ujedinjuje brojne informacije u sveobuhvatne PPI mreže za više od 1000 organizama (Jensen i sar. 2009; Szklarczyk i sar. 2011).

U današnje vreme, kompjuterska interakromika i predviđanje PPI se zasnivaju na više različitih vrsta informacija, tj. postoji tendencija da se predikcije donose na osnovu kombinacije podataka o fizičko-hemijskim karakteristikama aminokiselina, anotaciji proteina i proteinskih domena (npr. u GO, Pfam), podataka zasnovanih na znanju (npr. da li neki proteinski par interaguje u drugim vrstama), itd. Jedna grupa metoda vrši predviđanje realnih proteinskih kompleksa na osnovu modelovanja PPI mreža (Bader i Hogue 2003; Spirin i Mirny 2003; King i sar. 2004; Pereira-Leal i sar. 2004; Adamcsek i sar. 2006; Altaf-Ul-Amin i sar. 2006; Li XL i sar. 2007; Leung HC i sar. 2009; Wu M i sar. 2009). PPI mreža se predstavlja u vidu grafa, gde čvorovi (engl. *nodes*) predstavljaju proteine, a ivice (engl. *edges*) predstavljaju interakcije. Za identifikaciju kompleksa se primenjuju

principi teorije grafova, pri čemu proteinski kompleksi odgovaraju zbijenim subgrafovima u okviru PPI mreže. To znači da proteini koji su bliži u PPI mreži sa većom verovatnoćom formiraju realne komplekse. Razvijena je i grupa metoda za predikciju PPI, koje koriste statističke metode mašinskog učenja za pravljenje klasifikatora na osnovu primarne sekvence proteina i već postojećih podataka o interakcijama. Ove metode predstavljaju sekvence proteina u interagujućem paru kao vektore, koji se sastoje iz vrednosti posmatrane osobine za svaku aminokiselinu u sekvenci. Na primer, metod Bock i Gough se zasniva na naelektrisanju, hidrofobnosti i površinskom naponu aminokiselina (Bock i Gough 2001); metoda Liu i sar. se zasniva na PseAA kompoziciji analiziranih sekvenci (Liu L i sar. 2009); Shen i sar. su koristili elektrostatička i hidrofobna svojstva aminokiselina u vidu dipola i zapremine bočnog lanca (Shen J i sar. 2007); Ben-Hur i Noble su sa ovim ciljem kombinovali podatke o proteinskim sekvencama, anotacijama u GO, lokalnim svojstvima PPI mreža i homologim interakcijama u drugim vrstama (Ben-Hur i Noble 2005); itd. Ove metode treniraju svoje klasifikatore na pozitivnim i negativnim trening setovima, koristeći statističke metode poput *support vector machines*, *kernel functions*, *k-nearest neighbors*, ili njihove kombinacije.

S druge strane, predložene su kompleksne procedure za predikciju PPI, koje u prvom koraku preprocesovanja obuhvataju pretraživanje velikih baza proteina nekom od bioinformatičkih metoda opisanih ranije u ovom tekstu, nakon čega sledi filtriranje ovako dobijenih PPI parova, čime se smanjuje broj lažno pozitivnih rezultata. Mahdavi i Lin (Mahdavi i Lin 2007) su koristili dva tipa filtera: funkcionalnu anotaciju u GO i pravila zasnovana na *textmining* tehnikama. Roslan i saradnici (Roslan i sar. 2010) su, s druge strane, koristili tri vrste filtera: i) funkcionalnu anotaciju u GO ili dobijenu PFP procedurom (engl. *Protein Function Prediction*), ii) podatke o homologim interakcijama u drugim vrstama i iii) podatke o interagujućim domenima.

1.3.3. Bioinformatičke metode za predviđanje funkcionalnih domena proteina

Najveći broj bioinformatičkih alata za predikciju funkcionalnih domena proteina je zasnovano na MSA. MSA predstavlja uređeni niz sekvenci tako da su aminokisleniske rezidue u određenoj koloni homologe (poreklom iz iste pozicije u predačkoj sekvenci), da se preklapaju u rigidnom lokalnom strukturnom poravnanju i/ili da imaju zajedničku funkcionalnu ulogu (Edgar i Batzoglou 2006). Postoje različiti algoritmi za selekciju homologih sekvenci, koje će biti uključene u MSA, ali se svi zasnivaju na BLAST ili PSI-BLAST algoritmu za pretraživanje baza sekvenci (Altschul i sar. 1997). Najšire korišćeni alati za predikciju funkcionalnih domena proteina, kao što su Pfam (Finn i sar. 2014) i SMART (Letunic i sar. 2012), pored MSA, koriste i profile skrivenih Markovljevih modela (engl. *hidden Markov models*), koji omogućavaju poboljšanje MSA u slučaju postojanja insercija/delecija. Veliki broj funkcionalnih domena je uskladišten i u CDD bazi (engl. *Conserved Domains Database*), koja integriše podatke iz više različitih izvora (Marchler-Bauer i sar. 2013).

Postoje brojni dokazi da kod mnogih proteina postoje konzervirane, strukturno i funkcionalno značajne, periodičnosti koje ne odgovaraju konkretnoj aminokiselinskoj sekvenci, već su asocirane sa određenim fizičko-hemijskim karakteristikama aminokiselina. Razvijene su metode koje koriste vrednosti odgovarajućih fizičko-hemijskih parametara za numeričku reprezentaciju aminokiselina i detekciju latentnih obrazaca u proteinskoj sekvenci. Hidrofobnost je često korišćeni parametar za detekciju proteinskih regiona uključenih u savijanje – 'folding' (Eisenberg i sar. 1984). Identifikacija specifičnih obrazaca hidrofobnosti, detektovanih Furijeovom transformacijom, je iskorišćena za detekciju strukturnih domena GPCR proteina (Guo YZ i sar. 2006; Illingworth i sar. 2008). Marsella i saradnici su kreirali web alat REPETITA za detekciju solenoidnih ponovaka u proteinima (Marsella i sar. 2009). Ovi proteinski regioni se ne mogu detektovati pomoću MSA, već se s tim ciljem koriste metode zasnovane na Furijeovoj transformaciji. EIIP, na kom je zasnovana ISM metoda, se koristi kao parametar za detekciju funkcionalno značajnih

domena proteina (Cosic i sar. 1989; Cosic i Hearn 1991; de Trad i sar. 2000; Murakami 2000; Ivanović i sar. 2004; Sauren i sar. 2004; Veljkovic V i sar., 2009 (a), 2009 (b)).

1.3.4. Bioinformatičke metode za predviđanje funkcionalnih efekata AAS

Postoji nekoliko pristupa *in silico* predikciji funkcionalnih efekata polimorfizama, a najčešće korišćeni alati su PolyPhen-2 (Adzhubei i sar. 2010, 2013) i SIFT (Ng i Henikoff 2003; Kumar i sar. 2009). Većina metoda se zasnivaju na evolucionim informacijama dobijenim višestrukim poravnanjima sekvenci – MSA. Dve pretpostavke leže u osnovi evolucionog pristupa (Jordan i sar. 2010) su: i) varijante koje narušavaju biohemijske funkcije proteina ili izazivaju nepogodan fenotip, takođe, izazivaju i gubitak evolucionog fitnesa, što ih čini štetnim u evolucionom smislu; takve varijante se eliminišu iz populacije selekcijom i ii) štetne varijante u humanoj populaciji su, takođe, štetne i u homologim genima u drugim vrstama, tj. da je predeo evolucionog fitnesa konstantan. Funkcionalni značaj AAS predviđa se na osnovu obrasca aminokiselina koje se nalaze na određenoj poziciji u višestrukome poravnanju srodnih sekvenci. Varijanta koja ne odgovara ovom obrascu, smatra se štetnom, naročito ako je na visoko konzerviranoj poziciji. Evolucione metode obično imaju dva koraka. Prvi obuhvata odabir odgovarajućih homologih sekvenci i konstrukciju višestrukog poravnanja. Odabir sekvenci je kritičan jer su jako plitka poravnanja neinformativna, dok duboka poravnanja uključuju i udaljene sekvence, koje mogu da dovedu do pogrešnih zaključaka. Drugi korak je procena koliko alelska varijanta odgovara obrascu aminokiselina zapaženom u MSA, pomoću mera konzerviranosti pozicije, funkcija za skorovanje verovatnoće ili oba. Metode koje za određivanje evolucionih srodnosti sekvenci koriste *sequence weights* su: SIFT, PMut (Ferrer-Costa i sar. 2005) i MAPP (Stone i Sidow 2005). PolyPhen-2, s druge strane, koristi heuristički algoritam PSIC (engl. *Position-Specific Independent Counts*), što je slučaj i sa SNAP alatom (Bromberg i Rost 2007), na primer.

Pored evolucionih podataka, neke metode koriste dodatne informacije o strukturi, anotacijama i fizičko-hemijskim osobinama sekvenci divljeg tipa i mutiranog alela.

PolyPhen-2 koristi tri dodatna strukturna svojstva proteina, u situacijama kada protein ima poznatu 3D strukturu: zona aminokiseline divljeg tipa koja je dostupna na površini, promena sklonosti ka hidrofobnim regionima i kristalografski B-faktor, koji predstavlja konformacionu pokretnost aminokiseline divljeg tipa. Pored ovoga, PolyPhen-2 za svoje predikcije uzima u obzir i dubinu poravnjanja na mestu mutacije, promenu zapremine aminokiseline između divljeg tipa i mutiranog alela, a, takođe, i da li je pozicija aminokiseline u okviru anotiranog domena u Pfam bazi podataka. Ovaj alat je razvijen 2010. godine i najšire je korišćen alat za predikciju funkcionalnih efekata AAS, sa oko 2500 citata u naučnim radovima (u Google Scholar pretraživaču, pristupljeno 24.09.2014.). PolyPhen-2 je druga verzija PolyPhen alata (Ramensky i sar. 2002), koja je razvijena 2002. godine, pri čemu je i prva verzija bila u širokoj upotrebi, sa više od 1700 citata.

Mutacije koje se nalaze u konzerviranim regionima proteina najčešće narušavaju njegovu funkcionalnost (npr. Bienengraeber i sar. 2004; Scott i sar. 2004; Chan i sar. 2005). Zbog toga je predikcija funkcionalnih efekata AAS koje se nalaze van konzerviranih funkcionalnih domena (nCFD) u prošlosti bila zanemarena. Međutim, novija istraživanja pokazuju da se distribucija mutacija u kompleksnim bolestima, ne može razlikovati od distribucije SNP-ova (Kumar i sar. 2011). Takođe, više od 50% mutacija asociranih sa različitim tipovima kancera se nalazi u nCFD regionima proteina (Yue i sar. 2010; Peterson i sar. 2012). Analiza mutacija u važnim proteinskim porodicama asociranim sa patogenezom kancera, kao što je porodica proteinskih kinaza, je pokazala da se određeni broj 'driver' mutacija ne nalazi u kinaznim domenima, tj. nalaze se u nCFD (Stephens i sar. 2005; Greenman i sar. 2007). Zbog toga je razvoj nekih alata promenio fokus sa detekcije CFD mutacija, važnih za monogenska oboljenja, na detekciju i CFD i nCFD mutacija, uključenih u patogenezu kompleksnih bolesti (Torkamani i Schork 2007; Torkamani i sar. 2009; Shihab i sar. 2013 (a), 2013 (b)).

Analiza proteinskih sekvenci pomoću Furijeove transformacije ima prednosti u odnosu na metode zasnovane na MSA jer omogućava identifikaciju proteinskih karakteristika koje nisu evoluciono konzervirane u sekvenci, ali su neophodne za funkcionisanje proteina.

Takođe, prednost je korišćenje samo jedne sekvence za analizu, odnosno nema potrebe za postojanjem ni odabirom homologih sekvenci, što su kritični koraci u MSA analizi.

PseAAC (Shen HB i Chou 2008) je jedan od alata koji su zasnovani na analizi obrazaca fizičko-hemijskih osobina aminokiselina, kao što su hidrofobnost, hidrofilnost, veličina bočnog lanca, pK i pI, sa širokom primenom u predviđanju strukturnih klasa proteina, subćelijske lokalizacije ili mesta post-translacionih modifikacija proteina (Chou 2000, 2001, 2005; Chou i Cai, 2005; Shen HB i Chou, 2006). Međutim, do sada nije primenjen za predikciju funkcionalnih efekata polimorfizama. Drugi često korišćen alat iz ove kategorije je ISM, koji se zasniva na analizi obrazaca EIIP i predhodno je uspešno korišćen za funkcionalnu anotaciju AAS u humanim proteinima, uključujući važan tumor supresor TP53 (Veljkovic N i Perovic 2006; Glisic i sar. 2008).

2. Ciljevi istraživanja

Godišnja incidenca mijeloidnih maligniteta u Evropi je 7,5-8,6/100 000, a prosečna petogodišnja stopa preživljavanja pacijenata, uz primenu standardne terapije, je 37%. Zbog toga je neophodan razvoj novih terapijskih pristupa. Personalizovana medicina označava princip postizanja maksimalne terapijske efikasnosti kroz upotrebu ciljanih terapeutika kod biološki okarakterisanih pacijenata. Bioinformatika ima primenu u identifikaciji ciljnih molekula za razvoj terapije, racionalnom dizajnu lekova i identifikaciji informativnih biomarkera za prateću dijagnostiku. S obzirom da su proteini WT1, NISCH, ASXL1, DNMT3A, EZH2 i TET2 uključeni u patogenezu mijeloidnih maligniteta i da predstavljaju ciljne molekule za terapiju i/ili biomarkere, postavljeni su sledeći ciljevi ovog istraživanja:

1. Predviđanje novih PPI proteina WT1, koje su važne za patogeni efekat WT1 u mijeloidnim malignitetima.
2. Identifikacija i karakterizacija funkcionalnih domena proteina NISCH, koji mogu biti targeti za ciljanu terapiju mijeloidnih maligniteta.
3. Razvoj algoritma za predviđanje funkcionalnih efekata AAS u epigenetičkim regulatorima ASXL1, DNMT3A, EZH2 i TET2.

3. Materijal i metode

3.1. Podaci

3.1.1. Interaktori proteina WT1

Skup poznatih interaktora WT1 formiran je pretraživanjem literature, zaključno sa 31.12.2011. Selektovani su proteini za koje je eksperimentalno pokazano da fizički interaguju sa proteinom WT1.

3.1.2. AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2

Za kolekciju podataka o AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2, korišćena su tri izvora: literatura, baza somatskih mutacija u kanceru – COSMIC i baza SNP-ova – dbSNP. Literatura je pretraživana u PubMed pretraživaču, a odabrana su samo ona istraživanja koja su sadržala eksperimentalno testiranje patogenog statusa AAS. Patogeni status je smatran mutacijom ukoliko je pokazano da je AAS stečen u toku razvoja mijeloidnog maligniteta (somatska mutacija), tj. ukoliko je prisutan u obolelim ćelijama, ali ne i u drugim tkivima istog pacijenta (pr. koža, limfociti, itd.). Za AAS je označeno da ima neutralan patogeni status (SNP) ukoliko je pokazano njegovo prisustvo i u obolelom i u nekom drugom tkivu pacijenta, ili ukoliko je utvrđena učestalost pojavljivanja tog polimorfizma u zdravoj populaciji. COSMIC (Forbes i sar. 2008, 2011) je baza informacija, Wellcome Trust Sanger Instituta, o somatskim mutacijama i srodnim podacima asociranim sa humanim kancerima. S obzirom da za mnoge mutacije u COSMIC bazi ne postoje definitivni dokazi da su stečene, mutacije su preuzimane iz ove baze samo u slučaju potvrđenog somatskog porekla: '*Somatic Status*' = '*Confirmed Somatic*'. Izvor podataka o SNP-ovima, pored literature, bila je i baza Nacionalnog Centra za Biotehnoške

Informacije, dbSNP (Sherry i sar. 2001). SNP-ovi su preuzimani iz ove baze samo kada je definisana učestalost pojavljivanja polimorfizma u opštoj populaciji.

3.2. Metoda informacionog spektra – ISM

3.2.1. Osnove ISM metode

ISM je MSA-nezavisna metoda za analizu proteinskih sekvenci, koja se zasniva na brojanom kodiranju aminokiselina i Furijeovoj transformaciji tako dobijenog niza brojeva. Ova metoda se, u osnovi, sastoji iz dva koraka:

- i) Kodiranje aminokiselina brojevima koji predstavljaju vrednosti nekog fizičko-hemijskog parametra. Koristi se parametar EIIP, sa vrednostima datim u Tabeli 1.
- ii) Transformacija dobijenog niza brojeva u informacioni spektar, primenom Furijeove transformacije.

Alternativno, procedura može imati i treći korak, koji omogućava identifikaciju zajedničkih frekventnih komponenti informacionih spektara dva ili više proteina:

- iii) Konsenzus-spektralna analiza, tj. množenje svih informacionih spektara i generisanje konsenzus informacionog spektra (engl. *Consensus Informational Spectrum*, CIS). Množenjem dva informaciona spektra se generiše kros-spektar.

Algoritam je shematski predstavljen na [Slici 5](#).

EIIP predstavlja elektrodinamičko svojstvo aminokiselina koje, zajedno sa hidrofobnošću kao elektrostatičkim svojstvom, određuje osobine dalekosežnih interakcija molekula. EIIP se nalazi u bazi fizičko-hemijskih parametara, koji opisuju aminokiseline, AAindex (Kawashima i sar. 2008). Velikim brojem nezavisnih studija je pokazan njegov značaj za određivanje bioloških funkcija i međusobnih interakcija bioloških molekula (Ćosić i sar. 1986; Skerl i Pavlović 1988; Veljković V i Metlas 1988; Cosic i sar. 1989; Cosic i Hearn 1991; Cosic i sar. 1994; Krsmanovic i sar. 1998; de Trad i sar. 2000; Murakami 2000; Veljkovic N i sar. 2003; Ivanović i sar. 2004; Sauren i sar. 2004; Huang J i sar. 2005; Veljkovic N i Perovic 2006; Tintori i sar. 2007; Doliana i sar. 2008; Glisic i sar. 2008; Mancini i sar. 2009; Veljkovic V i sar. 2009 (a), 2009 (b); Tintori i sar. 2010; Mancini i sar. 2012; Stojković i sar. 2012; Perovic i sar. 2013).

EIIP predstavlja aproksimaciju energije valentnih elektrona u nekom molekulu i utiče na njegove dalekosežne interakcije jer određuje dipolni i kvadropolni momenat, jonizacioni potencijal, elektrofilnost i druga elektrodinamička svojstva tog molekula. Vrednosti EIIP za organske molekule se mogu izračunati na osnovu formule izvedene iz 'opšteg modela pseudopotencijala' (Veljković V i Slavić 1972), a vrednosti za 20 aminokiselina su date u Tabeli 1:

$$W = 0.25 \frac{Z^* \sin(1.04\pi Z^*)}{2\pi} \quad (1)$$

gde je Z^* srednji kvazivalentni broj, određen formulom:

$$Z^* = \frac{1}{N} \sum_{i=1}^m n_i Z_i \quad (2)$$

gde je Z_i valentni broj atoma, koji je i -ta komponenta molekula, n_i broj atoma i -te komponente, m broj komponenti molekula, a N ukupan broj atoma u molekulu.

Tabela 1. Vrednosti potencijala elektron jon interakcija (EIIP) za 20 aminokiselina.

Aminokiselina	Troslovni zapis	Jednoslovni zapis	EIIP vrednost
Alanin	Ala	A	0,03731
Arginin	Arg	R	0,09593
Asparagin	Asn	N	0,00359
Asparaginska kiselina	Asp	D	0,12630
Cistein	Cys	C	0,08292
Fenilalanin	Phe	F	0,09460
Glicin	Gly	G	0,00499
Glutamin	Gln	Q	0,07606
Glutaminska kiselina	Glu	E	0,00580
Histidin	His	H	0,02415
Izoleucin	Ile	I	0,00000
Leucin	Leu	L	0,00000
Lizin	Lys	K	0,03718
Metionin	Met	M	0,08226
Prolin	Pro	P	0,01979
Serin	Ser	S	0,08292
Tirozin	Tyr	Y	0,05159
Treonin	Thr	T	0,09408
Triptofan	Trp	W	0,05481
Valin	Val	V	0,00569

Furijeova transformacija predstavlja osnovnu analitičku metodu digitalne obrade signala. Ona omogućava transformaciju prostornog (ili vremenskog) domena nekog diskretnog signala u niz frekvencijskih domena, pri čemu se kao rezultat dobija spektar signala. U terminologiji ISM, gde se Furijeova transformacija upotrebljava za analizu proteina, ovaj spektar se naziva informacioni spektar. Kada se proteinska sekvenca predstavi nizom brojeva, Furijeova transformacija omogućava identifikaciju periodičnih strukturnih obrazaca u toj sekvenci. Diskretna Furijeova transformacija (DFT) se vrši tako što se diskretizuje jedna perioda Furijeove transformacije (koja je kontinualna funkcija) diskretnog signala, što omogućava obradu diskretnih signala na računaru. Brza Furijeova transformacija (engl. *Fast Fourier Transform*, FFT) predstavlja algoritam za vremenski

efikasno računanje DFT, koji se sastoji iz produžavanja ulaznog signala nulama (engl. *zero padding*) na dužinu jednaku najbližem većem broju koji predstavlja stepen broja dva. Kodirana proteinska sekvenca, tj. niz brojeva dobijen zamenom aminokiselina sa njihovim vrednostima EIIP, predstavlja diskretni signal, koji se konvertuje primenom Furijeove transformacije u spektar:

$$X(n) = \sum_{m=1}^N x(m) e^{-i2\pi(m-1)/N}, \quad n = 1, \dots, N/2 \quad (3)$$

gde je $x(m)$ m -ti član signala, $X(n)$ je n -ti koeficijent DFT, a N je dužina signala, tj. ukupan broj elemenata u EIIP numeričkom nizu, koji odgovara broju aminokiselina u proteinskoj sekvenci.

Koeficijenti DFT opisuju amplitudu/fazu i frekvencu sinusoida koje sačinjavaju originalni signal, a potpuna informacija o originalnoj sekvenci je sadržana u amplitudnom/faznom spektru. Za analizu proteina, relevantna informacija je sadržana u spektru gustine energije, definisanom:

$$S(n) = X(n)X^*(n) = |X(n)|^2, \quad n = 1, \dots, N/2 \quad (4)$$

S obzirom da je pretpostavka da se EIIP niz, kao ulazni signal, sastoji iz ekvidistantnih tačaka sa ukupnom distancom $d = 1$ i da je spektar signala realnih brojeva simetričan, maksimalna frekvencija u spektru je $F = 1/2d = 0,5$. To znači da je raspon frekvenci u spektru nezavisan od dužine proteinske sekvence. Dužina sekvence utiče na rezoluciju spektra: ukupna rezolucija spektra dužine N je $f = 1/N$, gde n -ti element spektra odgovara frekvenci $f(n) = nf = n/N$.

Na ovaj način se konstruiše informacioni spektar proteinske sekvence, koji je predstavljen nizom frekvenci i njihovih amplituda. Lokalni maksimumi spektra (pikovi) u informacionom spektru odgovaraju periodičnim strukturnim obrascima u proteinskoj sekvenci. S obzirom da se za konstrukciju informacionog spektra koristi EIIP, deskriptor dalekosežnih prepoznavanja bioloških molekula, zajednički frekvencijski domeni u

informativnim spektrima dva ili više proteina upućuju na njihovu specifičnu interakciju. PPI određuju biološke funkcije i procese u koje je neki protein uključen, pa se ISM analiza koristi za ispitivanje tih aspekata aktivnosti proteina.

Identifikaciju zajedničkih informativnih karakteristika dve ili više sekvenci omogućava CIS, koji se dobija konsenzus-spektralnom analizom. CIS za M spektara je definisan:

$$C(j) = \prod_{i=1}^M S(i, j), \quad j = 1, \dots, N/2 \quad (5)$$

gde je $S(i, j)$ j -ti element i -tog spektra, a $C(j)$ je j -ti element CIS. Tako definisan CIS je Furijeova transformacija konvolucije spektara. Na taj način se eliminiše svaka frekvencija koja nije prisutna u svim informativnim spektrima iskorišćenim za konstrukciju CIS. Lokalni maksimumi u CIS predstavljaju zajedničke frekventne komponente analiziranih sekvenci. Mera specifičnosti nekog pika u CIS je određena odnosom intenziteta tog pika na odgovarajućoj frekvenciji i srednje vrednosti na celom spektru (engl. *Signal to Noise ratio*, S/N).

Detekcija jasno definisanog spektralnog pika (ili malog broja pikova) u CIS dva ili više proteina upućuje na njihovu međusobnu interakciju i/ili zajedničku biološku funkciju ili proces u koji su ti proteini uključeni.

3.2.2. Programi EIIP/ISM platforme

Za različite analize su korišćeni programi softverskog paketa zasnovanog na ISM metodi i EIIP parametru. Ova platforma je razvijena u Centru za multidisciplinarna istraživanja INN Vinča (Perović 2013).

3.2.2.1. ProteinSpektar

ProteinSpektar je deo osnovnog modula ISM platforme i služi za generisanje informacionog spektra. Ovaj program računa spektre i DFT i FFT metodom i omogućava njihovo poređenje.

3.2.2.2. KrosSpektar

KrosSpektar predstavlja program za generisanje CIS. Ovaj program omogućava računanje CIS korišćenjem pojedinačnih sekvenci kao ulaz. Takođe, moguće je za neki protein konstruisati kros-spektre sa svakim pojedinačnim proteinom iz neke baze ili zajednički CIS za sve proteine u bazi. Konstrukcija CIS je moguća korišćenjem isključivo FFT metode. Korišćeni parametri za konstrukciju kros-spektara, tj. množenje amplituda uključenih spektara, proteina WT1 i pojedinačno svih humanih proteina iz UniProt, su:

- i) $\text{deltafrek} = 0,004$ (podrazumevana vrednost); ovaj parametar omogućava da se, pored amplitude na određenoj frekvenci, uzimaju u obzir i amplitude susednih frekvenci kao činiooci u množenju, pri čemu deltafrek određuje udaljenost tih susednih frekvenci;
- ii) metoda množenja amplituda = *Max&fiksirana prva*; ovaj način množenja amplituda fiksira spektar prve sekvence koja se unese u program (WT1), tj. za svaku frekvencu se gleda samo amplituda na centralnoj frekvenci ($\text{deltafrek} = 0$), a za ostale sekvence se za svaku frekvencu računa maksimalna amplituda u određenom frekventnom intervalu ($\text{deltafrek} = 0,004$).

3.2.2.3. AKSkener

AKSkener omogućava identifikaciju pojedinačnih aminokiselina ili regiona u sekvenci, koji najviše doprinose intenzitetu lokalnog maksimuma na određenoj frekvenci. Ovaj program funkcioniše tako što iseca N uzastopnih aminokiselina iz proteinske sekvence ili ih zamenjuje nekom drugom aminokiselinom. Iterativno se generiše skup novih sekvenci tako što se za poziciju od koje počinje isecanje/zamena uzimaju redom sve pozicije u ulaznoj

sekvenci od prve do poslednje moguće. *AKSkener* je zasnovan na FFT metodi i za svaku novodobijenu sekvencu računa ISM spektar i vrednosti amplitude i S/N na zadatoj frekvenci. *AKSkener* je korišćen u delu istraživanja koje se bavi kreiranjem procedure za predikciju funkcionalnih efekata AAS u epigenetičkim regulatorima. Cilj ove analize je bio utvrđivanje nezavisnosti rezultata ISM algoritma od pozicije AAS u CFD/nCFD, pa su parametri podešeni za virtuelno Ala skeniranje, tj. dužina prozora $N=1$, a aminokiseline su *zamenjene alaninom*. Statistička značajnost asociranosti rezultata virtuelnog Ala skeniranja i CFD/nCFD pozicija aminokiselina je analizirana Fišerovim testom. Pored toga, *AKSkener* je korišćen i za identifikaciju domena proteina NISCH, koji najviše doprinose amplitudama na značajnim frekvencama. U tom slučaju su korišćeni parametri $N=55-60$, a aminokiseline su *isečene*.

3.2.2.4. Mutacije i MutacijeDFT

Programi *Mutacije* i *MutacijeDFT* omogućavaju procenu uticaja pojedinačnih mutacija ili njihove kombinacije na informacione karakteristike ispitivanog proteina. Razlika između ova dva programa je što *Mutacije* koriste FFT, a *MutacijeDFT* DFT metodu. Ovi programi računaju amplitudu na određenoj frekvenci za mutiranu sekvencu, kao i procenat amplitude mutirane sekvence u odnosu na sekvencu divljeg tipa na odabranoj frekvenci. *MutacijeDFT* je korišćen za računanje ISM skorova u delu istraživanja, koje se bavi kreiranjem procedure za predikciju funkcionalnih efekata AAS u epigenetičkim regulatorima.

3.2.2.5. FilterDFTFFTBase

FilterDFTFFTBase omogućava selekciju proteina koji u svom informacionom spektru među frekvencama sa maksimalnom amplitudom imaju frekvencu koja je od interesa. Broj analiziranih frekvenci sa maksimalnom amplitudom podešava korisnik i u ovom istraživanju *broj prvih pikova* = 2. Ovaj program omogućava poređenje informacionih spektara bez računanja CIS i zato je korišćen za identifikaciju proteina sa zajedničkim frekventnim karakteristikama sa proteinom NISCH u spektrima generisanim DFT

metodom. Za analizu je korišćen preselektovani parametar $\text{deltafrek} = 0,004$. Selekcija proteina se vrši iz baze proteina sa prethodno izračunatim frekvencama i odgovarajućim amplitudama za odabrani broj prvih pikova. Zato je kreirana baza svih humanih proteina iz UniProt (Magrane i sar. 2011).

3.3. Ontologija gena i *enrichment* analiza GO termina

3.3.1. Ontologija gena

GO projekat omogućava korišćenje kontrolisanih rečnika definisanih termina, koji reprezentuju svojstva genskih produkata (Gene Ontology Consortium 2000). U okviru ovog projekta su razvijena tri strukturisana rečnika, tj. ontologije: i) ćelijska komponenta – delovi ćelije ili njenog vanćelijskog okruženja, ii) molekularna funkcija - elementarna aktivnost genskog produkta na molekulskom nivou, kao što je npr. katalitička aktivnost, i iii) biološki proces – operacija ili skup molekulskih događaja sa definisanim početkom i krajem, koji su relevantni za funkcionisanje jedinica živih sistema, kao što su ćelije, tkiva, organi i organizmi. Struktura GO ontologije se predstavlja grafom, gde je svaki GO termin čvor, a veze između termina su ivice između čvorova. GO ontologije su hijerarhijski organizovane, pri čemu su 'potomački' termini više specijalizovani u odnosu na 'pređačke' termine.

GO ontologija bioloških procesa je korišćena za predikciju PPI proteina WT1, kao i za identifikaciju bioloških procesa koji su asocirani sa pojedinim domenima proteina NISCH. Termin u okviru ove ontologije 'opisuje seriju događaja do koje dovodi jedan ili više skupova molekulskih funkcija', tj. proteinske interakcije sa drugim proteinima (PPI), nukleinskim kiselinama i/ili malim hemijskim molekulima. Primeri termina su 'signalna transdukcija' i 'proces metabolizma pirimidina'.

3.3.2. *GO Slims* i *QuickGO*

GO Slims (Gene Ontology Consortium 2013) predstavljaju liste GO termina selektovane iz punog skupa termina dostupnih u GO projektu. One se koriste za generisanje fokusiranog pogleda na deo GO ili za pregled širih kategorija u koje spadaju selektovani GO termini. *GO Slims* su generisani u analizi bioloških procesa u koje su uključeni identifikovani domeni proteina NISCH. Za vizualizaciju dela GO drvceta odabranih *GO Slims* je korišćen program *QuickGO* (Binns i sar. 2009). Za konstrukciju grafika su korišćeni podrazumevani parametri.

3.3.3. *DAVID*

DAVID (engl. *Database for Annotation, Visualization and Integrated Discovery*) platforma obuhvata alate za funkcionalnu anotaciju velikih skupova gena (Huang W i sar. 2009 (a), 2009 (b)). Važna strategija za funkcionalnu karakterizaciju gena je *enrichment* analiza genskih anotacija, kao što su GO termini, i ona predstavlja jedan od najvažnijih *DAVID* alata. Svi alati za *enrichment* analizu genskih anotacija se zasnivaju na pretpostavci da ukoliko je neki biološki proces relevantan za predmet datog istraživanja, geni koji učestvuju u ovom procesu imaju veću šansu da budu detektovani u tom istraživanju. Na primer, ukoliko su 10% gena detektovanih u 'high throughput' istraživanju kinaze, nasuprot 1% gena u humanom genomu koji su kinaze (što predstavlja referentni skup gena), može se zaključiti, na osnovu nekog od metoda za statističku *enrichment* analizu, da kinaze igraju važnu ulogu u predmetu istraživanja.

DAVID kao statistički metod koristi modifikovan Fišerov test, tj. EASE (engl. *Expression Analysis Systematic Explorer*) skorove, koji predstavljaju konzervativnu aproksimaciju Fišerove verovatnoće (Hosack i sar. 2003). Prednost EASE skorova je u tome što daju veću težinu anotacijama sa kojima je, u analiziranom skupu, asociiran veći broj gena. Ovo se postiže time što se računa Fišerova verovatnoća za neku kategoriju na

skupu gena u kom je jedan gen iz date kategorije uklonjen. Tako se smanjuje značajnost skupova koji sadrže mali broj gena.

Za *DAVID* analizu je, kao ulazna informacija, korišćena lista UniProt pristupnih brojeva (engl. *accession numbers*) proteina. Kao referentni genom je korišćen humani genom koji je deo *DAVID* alata za funkcionalnu analizu: opcija *Background*. Od *DAVID* alata su korišćeni:

➤ *Functional Annotation Clustering*. Ovaj alat grupiše slične, redundantne anotacione sadržaje iz istih ili različitih izvora, kao što su GO, INTERPRO, KEGG, itd., u anotacione klustere. Ovaj algoritam smanjuje opterećenje redundantnosti i omogućuje fokusiranje biološke interpretacije. Za opciju *Classification Stringency* je korišćena vrednost *High*, a za ostale opcije su korišćene podrazumevane vrednosti:

- *Kappa Similarity* – predstavlja meru slaganja između dva skupa kategorizovanih podataka, u ovom slučaju dva skupa gena asociranih sa anotacionim terminima: *Similarity Term Overlap* = 3, *Similarity Treshold* = 0,85;
- *Classification: Initial Group Membership* = 3, *Final Group Membership* = 3, *Multiple Linkage Treshold* = 0,50;
- *Enrichment Tresholds: EASE* = 1,0.

Enrichment skor rangira biološku značajnost klastera gena na osnovu ukupnih EASE skorova za sve obogaćene anotacione termine. Algoritam prvo analizira listu gena, koja predstavlja ulaznu informaciju za *DAVID*, i računa skor za svaki obogaćeni anotacioni termin, a, u drugom koraku, računa geometrijsku srednju vrednost EASE skorova anotacionih termina u klasteru.

➤ *Gene_Ontology Enrichment*. Odabrana je opcija za *enrichment* analizu GO termina asociranih sa biološkim procesima – *GOTERM_BP_FAT*. Ova analiza rangira GO termine sa kojima je asocirana lista gena, koja je korišćena kao ulazna informacija za *DAVID*, na osnovu EASE skorova. U analizi su korišćene podrazumevane vrednosti za *Tresholds: Count* = 2, *EASE* = 0,1.

3.4. STRING

STRING (engl. *Search Tool for the Retrieval of Interacting Genes*) je baza podataka posvećena interakcijama proteina, fizičkim i funkcionalnim (Jensen i sar. 2009; Szklarczyk i sar. 2011). Ovo je meta-baza, koja integriše informacije iz različitih izvora, uključujući repozitorijume literature, eksperimentalnih podataka i kompjuterskih predikcija. Osnovna jedinica interakcije u STRING bazi je ‘funkcionalna asocijacija’, koja je definisana kao specifična i smisljena interakcija između dva proteina, koji zajedno doprinose istom funkcionalnom procesu. Cilj STRING baze je da predstavi sve moguće međuproteinske veze, a filtriranjem izlaza iz STRING-a, npr. selekcijom proteina koji su koeksprimirani u nekom ćelijskom kontekstu, mogu se izvući zaključci o realnim PPI. Sa ovim ciljem je korišćen u proceduri za predviđanje novih PPI proteina WT1.

STRING sarži podatke o PPI iz 1100 organizama. Najveći broj interakcija u STRING bazi su predviđene asocijacije. Međutim, to nije slučaj sa humanim proteinima – većina su poznate interakcije identifikovane *textmining* tehnikama (Szklarczyk i sar. 2011). STRING obuhvata nekoliko miliona interakcija humanih proteina. Veze između proteina se predstavljaju PPI mrežama, koje su, kao i GO ontologije (v. Poglavlje 3.3.1.) reprezentovane grafom, gde je svaki protein čvor, a veze između proteina su ivice između čvorova. Rezultat pretraživanja STRING baze je sama mreža, koja se može preuzeti sa servisa, i nekoliko dokumenata sa podacima o proteinima u dobijenoj mreži i njihovim interakcijama.

STRING baza se pretražuje unosom imena/sekvence jednog (*‘search by name’/‘search by protein sequence’*) ili većeg broja proteina (*‘multiple names’/‘multiple sequences’*). Rezultat pretraživanja je deo celokupne PPI mreže zadatog organizma, koji se nalazi oko ulaznog/ulaznih proteina. U analizi WT1, kao ulazna informacija je korišćen skup svih poznatih interaktora ovog proteina zajedno sa samim WT1 proteinom. Tako dobijena PPI mreža reprezentuje proteinski milje u kom WT1 obavlja svoje biološke funkcije.

Značajna opcija STRING servisa je računanje nivoa poverenja svih PPI u bazi. Da bi neka PPI postala deo mreže mora da ima skor poverenja veći od granične vrednosti. Skor poverenja predstavlja kombinovani skor od četiri tipa sub-skorova (von Mering i sar. 2005): i) verovatnoća da se dva proteina, za koje se pretpostavlja da interaguju, nađu u istom biološkom putu; kao referentni skup bioloških puteva i asociiranih proteina, STRING koristi KEGG bazu; ii) verovatnoća interakcije na osnovu ‘*high-throughput*’ eksperimentalnih podataka; iii) sub-skor zasnovan na učestalosti i distribuciji imena gena i njihovih sinonima (pruzetih iz SwissProt) u abstraktima uskladištenim u PubMed bazi literature; iv) sub-skor zasnovan na funkcionalnoj genomici, npr. koregulacija gena u različitim eksperimentalnim uslovima; ove podatke STRING uvozi iz ArrayProspector servisa (Jensen i sar. 2004).

Selektovan je visok nivo poverenja, *confidence score* $\geq 0,700$. Korišćene su sve podrazumevane opcije za predikcione metode: *Neighborhood*, *Gene Fusion*, *Co-occurrence*, *Co-expression*, *Experiments*, *Databases* i *Textmining*. Odabrano je da STRING izlaz, pored ulaznih proteina, sadrži još 500 proteina u rezultujućoj PPI mreži.

3.5. Analiza koekspresije

Koekspresija WT1 i potencijalnih novih interaktora je ispitivana u K562 ćelijskoj liniji, koja ima BCR-ABL hromozomski rearanžman i predstavlja model za CML. Za analizu ekspresije proteina u K562 ćelijskoj liniji, korišćeni su *microarray* podaci uskladišteni u GEO (engl. *Gene Expression Omnibus*) bazi podataka (Edgar i sar. 2002; Barrett i sar. 2013). GEO je pretraživan po ključnoj reči K562. Izdvojeno je 12 studija koje obuhvataju uzorke K562 ćelija u bazičnom, netretiranom stanju, a ukupan broj uzoraka je 41 (Tabela 2).

Tabela 2. Identifikatori 12 studija iz *Gene Expression Omnibus* baze, sa 41 uzorkom K562 ćelija u bazičnom, netretiranom stanju.

Studija	Uzorak	Studija	Uzorak
GDS3558	GSM296608	GDS3046	GSM34921
	GSM296609		GSM34922
GDS3487	GSM304303	GDS3045	GSM34923
	GSM304304		GSM34915
	GSM304479		GSM34916
	GSM304480	GDS3044	GSM34917
	GSM304481		GSM34909
	GSM304482	GDS3043	GSM34910
	GSM304483		GSM34911
	GSM304484	GDS3043	GSM34134
	GSM304486		GSM34140
	GSM304498		GSM34146
GDS3049	GSM34939	GDS3042	GSM34039
	GSM34940		GSM34040
	GSM34941		GSM34041
GDS3048	GSM34933	GDS2970	GSM219392
	GSM34934		GSM219393
	GSM34935		GSM219394
GDS3047	GSM34927	GDS826	GSM16523
	GSM34928		GSM16524
	GSM34929		

Testiranje ekspresije gena je vršeno pomoću MAS 5.0 detekcionog algoritma, koji je deo *BioConductor* paketa *Affy* (Gautier i sar. 2004; Gentleman i sar. 2004), u *R* programskom okruženju (R Core Team 2012). Detektovano je prisustvo(P)/marginalno prisustvo(M)/odsustvo(O) ekspresije gena pomoću informatičkog koda:

```
library(affy)
affy.data = ReadAffy()
data.mas5calls = mas5calls(affy.data)
data.mas5calls.calls = exprs(data.mas5calls)
```

```
write.table(data.mas5calls.calls, file="uzorak_mas5calls.txt",
quote=F, sep="\t")
```

Ovako dobijeni dokumenti sadrže P/M/O rezultate za sve cDNK probe na *microarray* čipu za svaki pojedinačni uzorak. Svaki gen može biti reprezentovan sa jednom ili više proba na čipu, a broj i naziv proba zavisi od korišćene *microarray* platforme. Studije iz kojih su analizirani uzorci su koristile jednu od platformi: *HG-U133A* ili *HG-U133_Plus_2*. Nazivi proba, kojima su reprezentovani analizirani geni na ove dve platforme, su preuzeti iz *GATE Explorer* baze podataka (Prieto i sar. 2008; Risueño i sar. 2010). Za gene koji su reprezentovani sa 2 ili više proba na *microarray* čipu, smatra se da je ekspresija P (odnosno M) ukoliko je bar na jednoj probi detektovana ekspresija P (odnosno M).

3.6. Alati za ispitivanje funkcionalnih efekata polimorfizama

3.6.1. PolyPhen-2

PolyPhen-2 (engl. *Polymorphism Phenotyping v2*) je alat za predviđanje funkcionalnih efekata AAS, koji zasniva svoje predikcije na 11 svojstava selektovanih pomoću metoda mašinskog učenja (engl. *machine learning*) (Adzhubei i sar. 2010). Većina ovih svojstava podrazumeva poređenje alela divljeg tipa (pređačkog, normalnog) i mutiranog alela. Tu spada osam svojstava zasnovanih na sekvenci proteina:

- i) PSIC skor alela divljeg tipa (dt); PSIC skor predstavlja verovatnoću da neka određena aminokiselina zauzima specifičnu poziciju u proteinskoj sekvenci, s obzirom na obrazac AAS zapažen u višestrukom poravnanju sekvenci (Sunyaev i sar. 1999); PSIC algoritam uzima u obzir srodnost homologih sekvenci i koristi prethodno izvedene verovatnoće za AAS iz AAS matrice BLOSUM62;
- ii) PSIC skor mutiranog alela; predstavlja razliku PSIC skorova dt i mutiranog alela;
- iii) Identičnost sekvenci mutiranog alela i najbližeg homologa, koji nosi bilo koju AAS na mestu mutacije;
- iv) Kongruentnost mutantnog alela sa višestrukim poravnanjem;

- v) CpG kontekst tranzicionih mutacija;
- vi) Dubina poravnanja na mestu mutacije;
- vii) Promena u zapremini aminokiseline između dt i mutantne aminokiseline;
- viii) Da li je mesto mutacije pozicionirano u anotiranom Pfam domenu;

i tri svojstva zasnovana na 3D strukturi proteina:

- i) Dostupna površina dt aminokiselinske rezidue;
- ii) Promena hidrofobne sklonosti;
- iii) Kristalografski B faktor.

PolyPhen-2 koristi Naïve Bayes kao klasifikacioni metod, čiji je rezultat skor verovatnoće da je neka AAS štetna. AAS se klasifikuje kao 'verovatno štetna' ukoliko je skor verovatnoće veći od 0,85, 'moguće štetna' ukoliko je skor verovatnoće veći od 0,15 i 'benigna' u ostalim slučajevima. S obzirom da su PolyPhen-2 predikcije poređene sa predikcijama dobijenim drugim algoritmima, koji daju binarne predikcije, korišćena je binarna klasifikacija i za PolyPhen-2 i to: AAS je klasifikovana kao štetna mutacija ukoliko je skor verovatnoće bio veći od 0,50 i kao neutralan SNP u ostalim slučajevima (kao u npr. van Blitterswijk i sar. 2012; Liu X i sar. 2013; Vemula i sar. 2013).

Korišćena je *Batch Query Data* opcija PolyPhen-2 alata, koja omogućava istovremenu analizu velikog broja AAS. Kao model za klasifikaciju je korišćen *HumDiv*, što je podrazumevana opcija i preporučena za identifikaciju mutacija uključenih u kompleksne fenotipe, od strane autora alata. Za ostale opcije su, takođe, korišćene podrazumevane vrednosti: *Genome Assembly* = GRCh37/hg19, *Transcripts* = Canonical i *Annotations* = Missense.

3.6.2. SIFT

SIFT (engl. *Sorting Intolerant From Tolerant*) je alat za predviđanje efekata AAS na proteinsku funkciju (Ng i Henikoff 2003; Kumar i sar. 2009). Ovaj alat zasniva svoje

predikcije na stepenu konzerviranosti aminokiselinskih rezidua u poravnanju srodnih proteinskoj sekvenci, prikupljenih PSI-BLAST algoritmom. SIFT platforma obuhvata baze pre-kalkuliranih predikcija funkcionalnih efekata AAS, kao i četiri alata koji se razlikuju po tipu ulazne informacije. Korišćen je alat *SIFT Sequence*, koji podrazumeva sve korake SIFT procedure, uključujući i PSI-BLAST, na pojedinačnoj proteinskoj sekvenci. Procedura koju koristi ovaj alat sastoji se iz četiri koraka:

- i) Potraga za sličnim sekvencama. Vršiti se PSI-BLAST pretraga baze podataka korišćenjem ulazne sekvence. Pretraživana baza je *UniRef90 2011 Apr*. Korišćene su podrazumevane vrednosti za parametre: *Number of iterations = 2*, *Expectation value = 0,0001*, *E-value threshold for inclusion in multipass model = 0,002*.
- ii) Odabir srodnih sekvenci. Sekvence koje su dobijene PSI-BLAST analizom se grupišu tako što svakoj grupi pripadaju sekvence koje su među sobom identične više od postavljene granične vrednosti. U ovom radu je korišćena podrazumevana vrednost od 90%. Za svaku grupu se pravi konsenzusna sekvenca tako što se za svaku poziciju odabira najčešća aminokiselina u toj grupi sekvenci. Zatim se ulazna sekvenca poravnava sa najboljom konsenzusnom sekvencom. Konzerviranost se računa za svaku poziciju u poravnanju, a zatim i srednja vrednost. Ukoliko je ova srednja vrednost konzerviranosti veća od zadate granične vrednosti, konsenzusna sekvenca se zadržava u poravnanju. Ovaj proces se ponavlja dok srednja vrednost konzerviranosti ne padne ispod granične vrednosti. Korišćena je podrazumevana srednja vrednost konzerviranosti od 3,00.
- iii) Preuzimanje poravnanja iz inicijalnih PSI-BLAST rezultata.
- iv) Računanje verovatnoće. Na svakoj poziciji poravnanja, svaka aminokiselina se pojavljuje sa učestalošću n_i . Korišćenjem n_i vrednosti, verovatnoće za aminokiseline se procenjuju u skladu sa *Dirichlet mixture* modelom (d_i). Konačna verovatnoća da se neka aminokiselina pojavi na određenoj poziciji u sekvenci (p_i) se računa kao ponderisani prosek n_i i d_i . n_i se ponderišu na osnovu broja sekvenci korišćenih za konstrukciju poravnanja, a d_i na osnovu eksponencijalne funkcije mere diverziteta, koja se računa kao: $Div = \sum rank_i * n_i$, gde je $rank_i$ rang aminokiseline u odnosu na

dt aminokiselinu, kada se rangiraju substitucionni skorovi za dt aminokiselinu korišćenjem BLOSUM62. Dobijene verovatnoće se normalizuju deljenjem sa maksimalnom vrednošću.

Poređenjem sa eksperimentalnim podacima, pokazano je da su AAS sa verovatnoćom manjom od 0,005 štetne mutacije, pa se ova vrednost koristi kao granična za predikcije. Dakle, SIFT skor verovatnoće je obrnuto proporcionalan sa štetnošću AAS. S obzirom da je u drugim korišćenim algoritmima situacija obrnuta, tj. skorovi su direktno proporcionalni štetnosti mutacija, SIFT skorovi su transformisani: $SIFTskor = 1 - SIFTskor_{org}$, gde je $SIFTskor_{org}$ skor originalno dobijen SIFT alatom. Na primer, $SIFTskor_{org}$ od 0,01, koji odgovara predikciji štetne mutacije, i $SIFTskor_{org}$ od 0,88, koji odgovara predikciji neutralnog SNP-a, su transformisani u $SIFTskor$ 0,99 i 0,12, redom, tako da veći skor odgovara mutaciji, a manji SNP-u.

3.7. Statistička analiza

Prilikom konstrukcije i testiranja algoritma za predikciju funkcionalnih efekata AAS u epigenetičkim regulatorima ASXL1, DNMT3A, EZH2 i TET2 su korišćene statističke metode za analizu unakrsnih tabela, ROC (engl. *Receiver Operating Characteristic*) kriva, korelacija i razlika između dve grupe nezavisnih promenljivih. Sve analize su rađene u *SPSS Statistics 17.0* softveru. Statistički značajnim su se smatrali rezultati sa nivoom značajnosti od $p \leq 0,05$.

3.7.1. Unakrsne tabele

Za procenu performanse alata za predikciju funkcionalnih efekata AAS su korišćena tri parametra: *tačnost*, *senzitivnost* i *specifičnost*. Oni se izračunavaju na osnovu broja tačno pozitivnih (TP), tačno negativnih (TN), lažno pozitivnih (LP) i lažno negativnih (LN) binarnih predikcija, koji se predstavljaju unakrsnim tabelama (Tabela 3).

Tabela 3. Primer unakrsne tabele, sa tačno pozitivnim, tačno negativnim, lažno pozitivnim i lažno negativnim predikcijama alata za predviđanje funkcionalnih efekata aminokiselinskih supstitucija.

		Predviđene MUT/SNP	
		MUT	SNP
Eksperimentalno potvrđene MUT/SNP	MUT	TP	LN
	SNP	LP	TN

MUT – mutacije; SNP – polimorfizmi pojedinačnih nukleotida; TP – tačno pozitivne; TN – tačno negativne; LP – lažno pozitivne; LN – lažno negativne predikcije.

Parametri performanse se izračunavaju:

$$tačnost = (TP + TN)/(TP + TN + LP + LN) \quad (6)$$

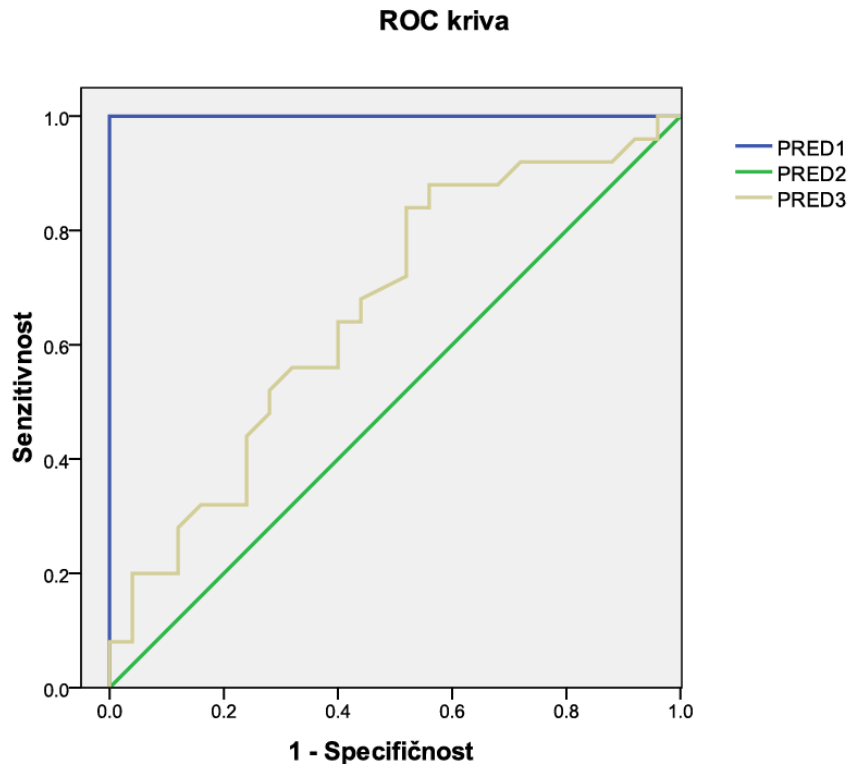
$$senzitivnost = TP/(TP + LN) \quad (7)$$

$$specifičnost = TN/(TN + LP) \quad (8)$$

Prilikom analize unakrsnih tabela, nivo statističke značajnosti asocijacije između dve kategoričke promenljive, kao što su predviđene i eksperimentalno potvrđene mutacije/SNP-ovi, testirao se Fišerovim testom.

3.7.2. ROC krive

Površina ispod ROC krive (engl. *Area Under the ROC Curve*, AUC) je mera performanse predikcionih alata, koja je bolja u odnosu na tačnost u situaciji kada alat daje, pored binarnih predikcija, i skorove verovatnoće ili neki drugi tip skorova (Bradley 1997; Huang J i Ling, 2005). ROC kriva se konstruiše na osnovu stope TP predikcija – na x osi je *senzitivnost*, i stope LP predikcija – na y osi je $(1-specifičnost)$ (Slika 6). Idealan predikcioni alat ima $AUC = 1$, dok alat koji daje nasumične predikcije ima AUC oko 0,500. Svi realni alati imaju AUC vrednost između ove dve (Slika 6). AUC je korišćena za poređenje alata za predikciju funkcionalnih efekata AAS.



Slika 6. Tri primera ROC krivih. PRED1 predstavlja primer performanse idealnog alata, sa $AUC = 1$; PRED2 predstavlja primer performanse alata koji daje nasumične predikcije, sa $AUC = 0,500$; PRED3 predstavlja primer performanse realnog alata, sa $AUC = 0,701$.

AUC – Area Under ROC Curve; ROC – Receiver Operating Characteristic.

3.7.3. Korelaciona analiza

Korelacija opisuje jačinu i smer linearne veze između dve promenljive. Koeficijenti korelacije mogu imati vrednosti od -1 do +1. Predznak pokazuje da li je korelacija pozitivna (obe promenljive zajedno i opadaju i rastu) ili negativna (jedna promenljiva opada kada druga raste, i obrnuto). Apsolutna vrednost tog koeficijenta pokazuje jačinu veze. Savršena korelacija, koja iznosi 1 ili -1, pokazuje da se vrednost jedne promenljive može tačno utvrditi kada znamo vrednost druge. S druge strane, korelacija jednaka nuli

pokazuje da između te dve promenljive ne postoji nikakva veza. Korišćene su sledeće smernice za jačinu korelacije (Pallant 2011):

Mala: $0,10 \geq \text{koef. korelacije} \leq 0,29$,

Srednja: $0,30 \geq \text{koef. korelacije} \leq 0,49$,

Velika: $0,50 \geq \text{koef. korelacije} \leq 1,00$.

U ovom radu je korišćena Spirmanova analiza korelacija. Spirmanov koeficijent korelacije (r_o) je prikladan za ordinalne veličine ili veličine koje se mogu rangirati, ima manje stroge kriterijume koje podaci moraju da zadovolje u odnosu na Pirsonovu korelaciju, pa se najčešće koristi kao neparametarska tehnika korelacione analize.

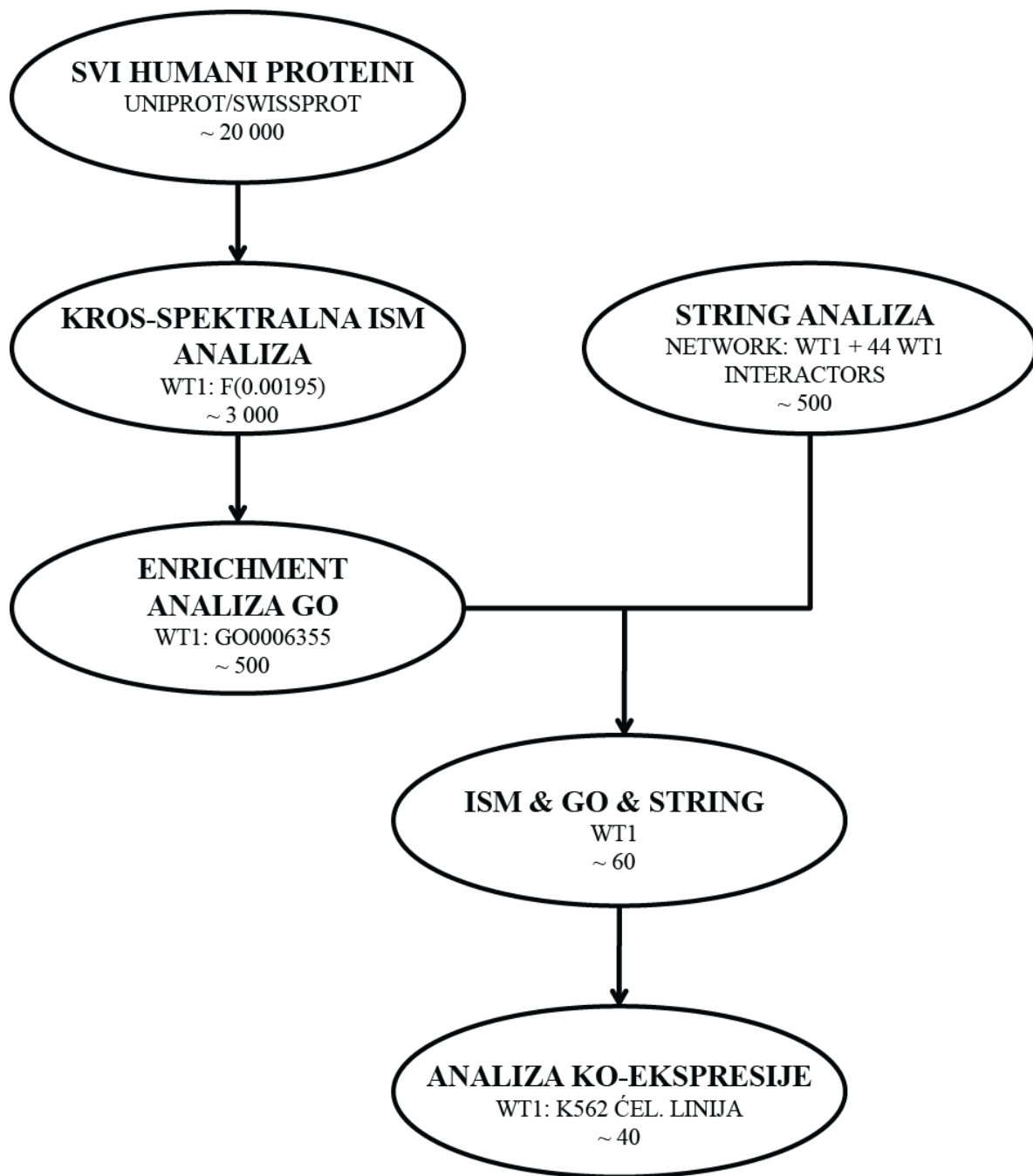
3.7.4. Mann-Whitney U Test

Mann-Whitney U test se upotrebljava za ispitivanje razlika između dve nezavisne grupe na neprekidnoj skali i predstavlja neparametarsku alternativu t-testu. Ovaj test pretvara vrednosti promenljive u rangove za obe grupe i potom izračunava da li se rangovi tih grupa značajno razlikuju.

3.8. Algoritam za predikciju PPI

Za predviđanje novih PPI proteina WT1 je korišćen integrativni algoritam prikazan na [Slici 7](#). Obuhvata pet koraka, a svaki je opisan u ranijim poglavljima:

- i) Kros-spektralna ISM analiza
- ii) *Enrichment* analiza GO
- iii) STRING analiza
- iv) Integrisanje rezultata ISM, GO i STRING analiza
- v) Analiza koekspresije



Slika 7. Integrativni algoritam za predikciju protein-protein interakcija.

GO – ontologija gena; ISM – metod informacionog spektra.

4. Rezultati

4.1. Predikcija novih PPI proteina WT1

4.1.1. Lista poznatih interaktora proteina WT1

Pregledom literature, identificirano je 42 proteina (zaključno sa 2011. godinom) koji direktno interaguju sa WT1 (Tabela 4).

Tabela 4. Proteinski partneri WT1 identificirani do kraja 2011. godine.

Funkcija PPI	UniProt ID	Interaktor	Alternativna imena	Literatura
Transkripcija	P04637	Cellular tumor antigen p53 (TP53)	P53	Maheswaran i sar. 1993
	Q96IZ0	PRKC apoptosis WT1 regulator protein (PAWR)	PAR4	Johnstone i sar. 1996
	O76071	Probable cytosolic iron-sulfur protein assembly protein CIAO1 (CIAO1)	WDR39	Johnstone i sar. 1998
	Q13285	Steroidogenic factor 1 (NR5A1)	AD4BP, FTZF1, SF1	Nachtigal i sar. 1998
	O94993	Transcription factor SOX-30 (SOX30)	/	Osaki et al. 1999
	Q00403	Transcription initiation factor IIB (GTF2B)	TF2B, TFIIB	McKay i sar. 1999
	Q9H3D4	Tumor protein 63 (TP63)	KET, P63, P73H, P73L, TP73L	Scharnhorst i sar. 2000
	O15350	Tumor protein p73 (TP73)	P73	Scharnhorst i sar. 2000
	Q15007	Pre-mRNA-splicing regulator WTAP (WTAP)	KIAA0105	Little i sar. 2000
	Q92793	CREB-binding protein (CREBBP)	CBP	Wang W i sar. 2001
	P20226	TATA-box-binding protein (TBP)	GTF2D1, TF2D, TFIID	Richard i sar. 2001
	Q14192	Four and a half LIM domains protein 2 (FHL2)	DRAL, SLIM3	Du i sar. 2002
	Q02962	Paired box protein Pax-2 (PAX2)	/	Discenza i sar. 2003
	Q05066	Sex-determining region Y protein	TDF	Matsuzawa-

		(SRY)		Watanabe i sar. 2003
	P80723	Brain acid soluble protein 1 (BASP1)	NAP22	Carpenter i sar. 2004
	P36956	Sterol regulatory element-binding protein 1 (SREBF1)	BHLHD1, SREBP1	Rae i sar. 2004
	Q12772	Sterol regulatory element-binding protein 2 (SREBF2)	BHLHD2, SREBP2	Rae i sar. 2004
	A6NIX2	Wilms tumor protein 1-interacting protein (WTIP)	/	Srichai i sar. 2004
	P03372	Estrogen receptor (ESR1)	ESR, NR3A1	Reizner i sar. 2005
	P40763	Signal transducer and activator of transcription 3 (STAT3)	APRF	Rong i sar. 2006
	Q99967	Cbp/p300-interacting transactivator 2 (CITED2)	MRG1	Val i sar. 2007
	P43694	Transcription factor GATA-4 (GATA4)	/	Miyamoto i sar. 2008
	Q5JTC6	APC membrane recruitment protein 1 (AMER1)	FAM123B, WTX	Rivera et al. 2009
	Q9NZL3	Zinc finger protein 224 (ZNF224)	BMZF2, KOX22, ZNF233, ZNF255, ZNF27	Florio i sar. 2010
	P26358	DNA (cytosine-5)-methyltransferase 1 (DNMT1)	AIM, CXXC9, DNMT	Xu B i sar. 2011
	Q15910	Histone-lysine N-methyltransferase EZH2 (EZH2)	KMT6	Xu B i sar. 2011
	O00255	Menin (MEN1)	SCG2	Xu B i sar. 2011
	Q15022	Polycomb protein SUZ12 (SUZ12)	CHET9, JAZ1, KIAA0160	Xu B i sar. 2011
	P48059	LIM and senescent cell antigen-like-containing domain protein 1 (LIMS1)	PINCH, PINCH1	Wang CY i sar. 2011
Obrada iRNK	P26368	Splicing factor U2AF 65 kDa subunit (U2AF2)	U2AF65	Davies i sar. 1998
	Q9NZL3	Bone marrow zinc finger 2 (BMZF2)	ZNF224 , KOX22, ZNF233, ZNF255, ZNF27	Lee TH i sar. 2002
	P23246	Splicing factor, proline- and glutamine-rich (SFPQ)	PSF	Morrison AA i sar. 2006
	Q9BWF3	RNA-binding protein 4 (RBM4)	RBM4A	Markus i sar. 2006
	Q00839	Heterogeneous nuclear ribonucleoprotein U (HNRNPU)	HNRPU, SAFA, U21.1	Spraggon i sar. 2007
PTM	P17612	cAMP-dependent protein kinase catalytic subunit alpha (PRKACA)	PKACA	Ye i sar. 1996
	P17252	Protein kinase C alpha type (PRKCA)	PKCA, PRKACA	Ye i sar. 1996

	P63279	SUMO-conjugating enzyme UBC9 (UBE2I)	UBC9, UBCE9	Wang ZY i sar. 1996
	O14933	Ubiquitin/ISG15-conjugating enzyme E2 L6 (UBE2L6)	UBCH8	Makki i sar. 2008
Drugo	P08107	Heat shock 70 kDa protein 1A/1B (HSPA1A/ HSPA1B)	HSPA1, HSX70	Maheswaran i sar. 1998
	P60709	Actin, cytoplasmic 1 (ACTB)	/	Dudnakova i sar. 2010
	P07900, P08238	Heat shock protein HSP 90-alpha (HSP90AA1), Heat shock protein HSP 90-beta (HSP90AB1)	HSP90A, HSPC1, HSPCA, HSP90B, HSPC2, HSPCB	Bansal i sar. 2010
	O43464	Serine protease HTRA2, mitochondrial (HTRA2)	OMI, PRSS25	Hartkamp i sar. 2010

PPI – protein-protein interakcija; PTM – post-translacione modifikacije.

Nakon 2011. godine, otkrivena su još četiri interaktora WT1: importini α i β (KPNA2 i KPNB1) (Depping i sar. 2012), SOX11 (Murugan i sar. 2012) i p65 subjedinica NF-kB (RELA) (Ristola i sar. 2012).

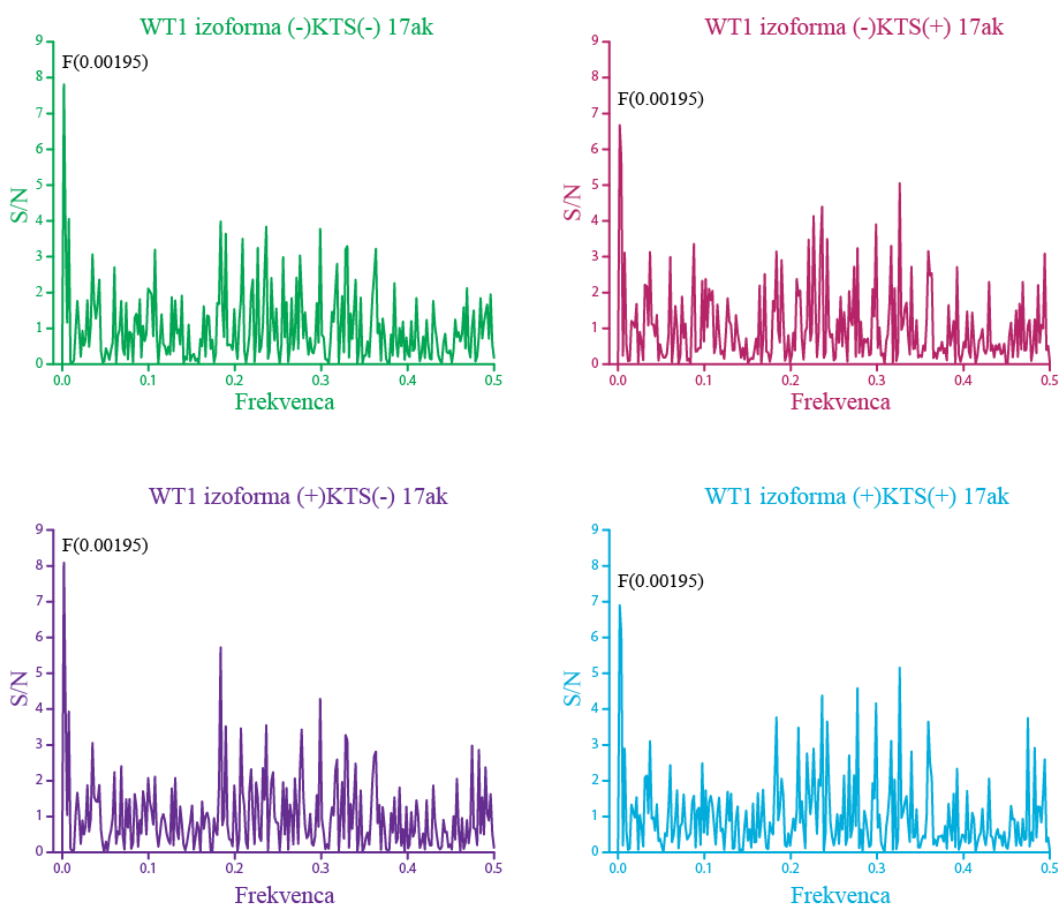
4.1.2. Predviđanje PPI proteina WT1

Predviđanje proteinskih interakcija WT1 vršili smo primenom integrativnog algoritma opisanog u Metodama.

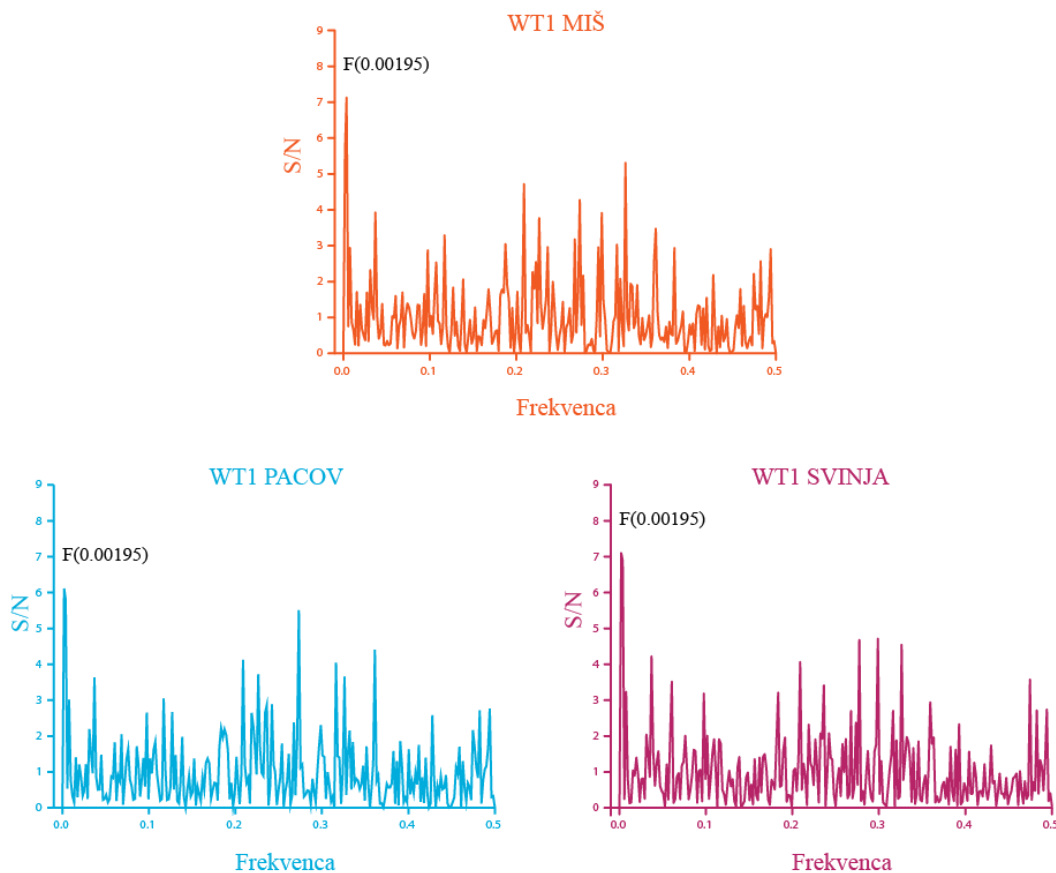
4.1.2.1. Kros-spektralna analiza WT1

Za identifikaciju najznačajnije frekvence u informacionom spektru proteina WT1, korišćen je program ISM platforme – *ProteinSpektar*, FFT metod. U spektru humane WT1 sekvence uočava se jedan dominantan lokalni maksimum na frekvenci F(0,00195) ([Slika 8](#)). Komparacija spektara svih dostupnih sekvenci WT1 sisara – miša, pacova i svinje ([Slika 9](#)), pokazala je da je ova frekvencija dominantna i u njihovim spektrima. Filogenetska konzerviranost dominantne osobine spektra ukazuje na njenu korelaciju sa najvažnijim funkcijama WT1 proteina.

Kao što je već rečeno, WT1 se može naći u četiri glavne izoforme. Različite izoforme WT1 učestvuju u različitim biološkim procesima, transkripciji ili obradi RNK, zavisno od prisustva KTS insercije između trećeg i četvrtog cink-prstića WT1 (Larsson i sar. 1995). Zato je testirano kako ova insercija utiče na svojstva informacionog spektra WT1. Kao što se vidi na [Slici 8](#), $F(0,00195)$ je dominantna karakteristika i WT1 (-)KTS izoformi, koje su najviše uključene u transkripciju, i WT1 (+)KTS izoformi, koje prevashodno vezuju RNK i učestvuju u splajsovanju. Na osnovu ovih rezultata, može se zaključiti da je frekvenca $F(0,00195)$ zaista najvažnija osobina spektra WT1, pa je ona odabrana za skrining humanog proteoma u potrazi za proteinima koji je imaju u svom spektru, a koji bi predstavljali potencijalne WT1 partnere.



Slika 8. Informacioni spektri četiri glavne izoforme humanog WT1. Dominantan je lokalni maksimum na frekvenci $F(0,00195)$.



Slika 9. Informacioni spektri sekvenci WT1 miša, pacova i svinje. Dominantan je lokalni maksimum na frekvenci F(0,00195).

U nastavku su u *KrosSpektar* programu generisani kros-spektri za WT1 i pojedinačno sve druge humane proteine iz UniProt baze. Izdvojeni su proteinski kandidati sa dominantnom frekvencom na F(0,00195). Meru specifičnosti nekog vrha u spektru predstavlja odnos signala i šuma – S/N, tj. odnos između intenziteta vrha na nekoj određenoj frekvenci i srednje vrednosti čitavog spektra. Kritični korak u ovom delu istraživanja je određivanje granične vrednosti S/N, koja je odgovarajuća za prepoznavanje biološki značajnih predikcija. Realne interakcije WT1 su korišćene za definisanje granične vrednosti S/N. Na osnovu Tabele 5, granična vrednost je S/N(0,00195) spektra proteina WT1 i njegovog interaktora serin proteaze HTRA2 i iznosi 10,00.

Tabela 5. S/N(0,00195) iz kros-spektara WT1 i njegovih poznatih interaktora, u kojima je F(0,00195) dominantna frekvenca.

		Protein	S/N(0,00195)
WT1 Izoforma (-)KTS(-)17ak	+	NR5A1	35,30
		U2AF2	25,16
		TBP	22,59
		ESR1	19,42
		SFPQ	18,56
		PAWR	16,68
		RBM4	15,56
		TP73	15,46
		SOX30	14,95
		BASP1	13,70
		TP63	11,09
		HTRA2	10,57

Lista proteinskih kandidata dobijenih kros-spektralnom analizom sadrži 2 576 kandidata – LISTA1 (Prilog 1).

Važno je naglasiti da LISTA1 sadrži i poznate kofaktore WT1 uključene u regulaciju transkripcije: TP63, TP73, TBP, ESR1, NR5A1 i BASP1, ili metabolizam RNK: PAWR, RBM4, SOX30, U2AF2 i SFPQ (Tabela 4). Takođe, u proteinskom skupu LISTA1 je i protein p65 (RELA), član NFkB kompleksa, čiju su interakciju sa WT1 otkrili Ristola i saradnici (Ristola i sar. 2012), nakon ove studije.

4.1.2.2. *Enrichment* analiza GO integrativnog algoritma primenjena na WT1

Da bi se identifikovali biološki procesi sa kojima je asocirana spektralna karakteristika WT1 proteina na F(0,00195), urađena je *enrichment* analiza GO termina, pomoću *Functional Annotation Clustering* DAVID platforme. Ulazna informacija je LISTA1, a referentni genom je celokupni humani genom. Ova analiza je pokazala značajan *enrichment* elementima uključenim u regulaciju transkripcije i obradu iRNK (Tabela 6).

Tabela 6. DAVID *enrichment* skorovi za funkcionalne klastere dobijene analizom skupa proteina LISTA1.

Anotacioni klaster	<i>Enrichment</i> skor
Regulacija transkripcije	20,77
Obrada iRNK	11,96
Motiv prepoznavanja RNK	10,64
Transkripcioni faktor ' <i>fork-head</i> '	9,48
Unutrašnjost organela	8,49

Iz Tabele 6 se vidi da je najvažniji biološki proces sa kojim je asociirana $F(0,00195)$, regulacija transkripcije. Da bi se identifikovale nove PPI proteina WT1, koje modulišu njegovu ulogu u ovom procesu, dalje predikcije su fokusirane na podskup skupa LISTA1, koji obuhvata proteine uključene u regulaciju transkripcije (LISTA2). LISTA2 sadrži 504 proteina (Prilog 2), koji su u GO anotirani terminom GO:0006355. Ovaj termin je definisan kao 'Bilo koji proces koji modulira frekvencu, stopu i obim ćelijske DNK-zavisne transkripcije'.

4.1.2.3. Mrežna analiza WT1

Mapiranje WT1 proteinskog okruženja izvršeno je STRING analizom. Iskorišćena je opcija unosa većeg broja proteina kao ulazne informacije za konstrukciju PPI mreže: pored WT1, ulaz su predstavljali i svih 43 poznata interaktora WT1 (v. Tabela 4). Konstruisana je mreža koja je, pored 44 uneta proteina, sadržala još 500 kandidata za interaktore WT1. Za konstrukciju mreže su odabrane PPI koje imaju visok skor poverenja. U finalnoj mreži najniži skor je bio 0,962. Za najveći broj PPI u dobijenoj mreži originalni izvori u STRING-u su eksperimentalni podaci, baze podataka i *textmining*.

Na ovaj način je dobijena lista od 500 proteina, LISTA3 (Tabela 7), koja će se u nastavku koristiti za integrisanje rezultata sa ISM i *enrichment* analizom GO termina.

Tabela 7. LISTA3 dobijena STRING analizom WT1 i njegovih poznatih interaktora.

ABCA1	CDK7	FKBP4	IKBKB	MTOR	POLR2H	S100B	TAF2E
ABL1	CDK8	FKBP5	IL17A	MYB	POLR2I	SAFB	TAF3
ACACB	CDK9	FOS	IL23A	MYC	POLR2L	SAP30BP	TAF4
ACTN2	CDKN1A	FOXA1	IL23R	MYOD1	POLR3C	SCAP	TAF5
AEBP2	CDKN2A	FOXO1	IL2RA	MYST1	POLR3F	SDC4	TAF7
AHR	CEBPB	FOXO4	IL2RB	NCBP2	POMC	SETD7	TAF8
AHSA1	CEBPD	GADD45A	IL6	NCK2	POU2F1	SF1	TAF9
AIRE	CFTR	GAPDH	IL6R	NCL	POU4F1	SF3A2	TBX5
AK6	CHEK1	GATA1	IL6ST	NCOA1	POU4F2	SF3B1	TCERG1
AKT1	CHEK2	GDF15	ILK	NCOA2	PPARG	SF3B2	TERT
APAF1	CHERP	GHR	ING1	NCOA3	PPARGC1A	SF3B4	TFAP2A
APC	CITED1	GJA1	ING4	NCOA6	PPP1R13B	SFN	TFF1
AQR	CKN2	GRB2	INSIG1	NCOA7	PPP1R13L	SIN3A	TGS1
AR	COPS5	GRIN2A	INSIG2	NCOR1	PPP5C	SIRT1	THAP1
ARID3A	CREB1	GRIN2B	IPO13	NCOR2	PRKAB1	SKP2	THRA
ARIH1	CRP	GRIP1	IRF1	NDRG1	PRKAR1A	SMAD1	THRB
ASXL1	CSF2	GSK3B	IRF3	NDUFA13	PRKCD	SMAD2	TMSB4X
ATF3	CSNK1D	GTF2A1	IRF7	NEDD8	PRKDC	SMAD3	TNFRSF10A
ATM	CSNK2A1	GTF2A2	IRF9	NFATC2	PRL	SMAD4	TNFRSF10B
ATR	CTBP1	GTF2E1	ISG15	NFATC4	PRMT1	SMARCA4	TNFRSF10D
AURKA	CTCF	GTF2E2	ITGB1	NFE2L2	PRMT5	SMYD2	TOP1
AXIN1	CTNNB1	GTF2F1	ITGB4	NFYA	PRPF3	SNAPC1	TOPORS
BAD	CTSD	GTF2F2	JAK1	NFYB	PRPF31	SNAPC2	TP53AIP1
BAG1	CUL9	GTF3C4	JAK2	NKX2-5	PTBP1	SNAPC4	TP53BP1
BAG3	CXCL12	GTF3C5	JAK3	NOC2L	PTBP2	SNIP1	TP53BP2
BAK1	CXCR4	HAND2	JUN	NONO	PTEN	SNRPA	TP53I3
BARD1	CXorf56	HCK	JUND	NOS3	PTGES3	SNRPC	TP53INP1
BAX	DAXX	HDAC1	KAT2B	NPAS2	PTGS2	SOCS3	TP53RK
BBC3	DDB2	HDAC2	KAT5	NPM1	PTPN1	SOD1	TPT1
BCAR1	DDX17	HDAC3	KDM1A	NR0B1	PTPN11	SOS1	TRAF6
BCL2	DDX5	HDAC4	KEAP1	NR0B2	PTPN2	SP1	TRIAP1
BCL2L1	DDX54	HDAC6	KHDRBS1	NR2C1	PUF60	SPI1	TRIM24
BCL6	DDX58	HELLS	KHDRBS3	NR2F1	QKI	SRC	TRIM25
BIRC2	DGCR14	HERC5	KLF4	NR3C1	RAC1	SRF	TRRAP
BIRC6	DHX9	HEY2	KPNB1	NRIP1	RAD51	SRSF1	TYK2
BLM	DMAP1	HIC1	KRAS	PABPC1	RAF1	SRSF11	U2AF1
BRAF	DNMT3A	HIF1A	LCK	PADI4	RANBP2	SRSF2	UBA2
BRCA1	DNMT3B	HIPK2	LCOR	PARP1	RANGAP1	SRSF2IP	UBC
BRCA2	DR1	HIST1H3A	LDLCQ3	PBK	RARA	SRSF3	UBE2K
BRD7	DRAP1	HK2	LEP	PCNA	RB1	STAT1	UBE3A
BRF1	DYRK2	HMGA1	LEPR	PDGFRB	RB1CC1	STAT2	UHRF1
BTAF1	E2F1	HNF1A	LIF	PELP1	RBBP4	STAT4	USP7
BTRC	E2F5	HNF4A	LMO4	PERP	RBBP7	STAT6	VAV1
C12orf5	E4F1	HNRNPA0	MAGEA2	PFN1	RBM10	STIP1	VDR
C16orf80	EBF1	HNRNPD	MAPK1	PHB2	RBM22	STK11	VEGFA

CARM1	EED	HNRNPF	MAPK11	PHF1	RBM23	STUB1	VRK1
CASC1	EGF	HNRNPH2	MAPK14	PIAS1	RBM39	SUGP1	WDR77
CBX4	EGFR	HNRNPH3	MAPK3	PIAS2	RBM42	SUGT1	WRN
CCAR1	EGR1	HNRNPK	MAPK8	PIAS3	RBM5	SUMO1	WWOX
CCNA2	EIF2AK2	HNRNPM	MAPK9	PIAS4	RBM7	SUMO2	XBP1
CCNB1	EP300	HNRNPUL1	MARCKS	PICK1	RBM9	SUMO3	XIAP
CCNC	EPHA2	HRAS	MBTPS1	PIK3CA	RCHY1	SUV39H1	XPO1
CCND1	ERBB2	HSF1	MDM2	PIK3R1	RELA	SYK	XRCC6
CCNG1	ERCC2	HSPA4	MDM4	PIN1	RFWD2	SYNCRIP	YAP1
CCR5	ERCC3	HSPA8	MECP2	PLD1	RICTOR	TADA3	YBX1
CD82	ESR2	HSPA9	MED1	PLD2	RING1	TAF1	YWHAZ
CDC25C	ESRRA	HSPD1	MED15	PLK3	RNF20	TAF10	YY1
CDC37	EZH1	HTT	MEF2C	PML	RPA1	TAF11	ZFPM1
CDC40	EZR	IFNAR1	MLH1	POLR2A	RPS6KA1	TAF12	ZFPM2
CDH2	FAM96B	IGF1	MLL	POLR2B	RRM2B	TAF13	
CDK1	FASN	IGF1R	MMS19	POLR2E	RXRA	TAF1A	
CDK2	FBXO11	IGFBP3	MSH2	POLR2F	S100A2	TAF1C	
CDK5	FDFT1	IGFBP5	MTA1	POLR2G	S100A4	TAF2	

4.1.2.4. Integriranje rezultata ISM, GO i STRING analize proteina WT1

Proteini zajednički za skupove LISTA2, dobijen sukcesivnim ISM i GO *enrichment* analizama, i LISTA3, dobijen STRING analizom (v. shemu na [Slici 7](#)), čine skup kandidata za nove interaktore WT1 (LISTA4). LISTA4 se sastoji od 51 proteina (Tabela 8), koji imaju sve sledeće karakteristike: i) dominantan lokalni maksimum na frekvenci $F(0,00195)$ u kros-spektru sa proteinom WT1, $S/N \geq 10$, ii) uključeni su u regulaciju transkripcije (anotirani GO terminom GO:0006355) i iii) nalaze se u najbližem susedstvu (koje uključuje 500 proteina) WT1 u STRING mreži humanih PPI.

Tabela 8. LISTA4 predstavlja skup kandidata za nove interaktore WT1. Prikazana je koekspresija proteina iz ovog skupa sa WT1 u K562 ćelijskoj liniji; koeksprimiranim se smatraju proteini prisutni (P+M) u $\geq 80\%$ uzoraka (obeleženi zelenom bojom).

Gen	P (%)	M (%)	O (%)	Gen	P (%)	M (%)	O (%)
AEBP2	100	0	0	MED15	100	0	0
AR	51	15	34	MEF2C	100	0	0
CDK2	100	0	0	MTA1	100	0	0
CDK8	100	0	0	NFATC2	100	0	0
CDK9	86	7	7	NFATC4	10	0	90
CTCF	100	0	0	NKX2-5	0	0	100
CTNNB1	100	0	0	POU2F1	88	7	5
DDX54	100	0	0	PPARGC1A	0	0	100
DMAP1	100	0	0	PPP1R13L	37	10	53
DNMT3B	100	0	0	PUF60	100	0	0
EGR1	100	0	0	RBM39	100	0	0
FOS	12	2	86	RING1	100	0	0
FOXO1	41	5	54	SAFB	100	0	0
FOXO4	51	5	44	SF1	100	0	0
GTF2E1	100	0	0	SNIP1	98	0	2
HNF1A	5	7	88	SRF	73	7	20
HNF4A	2	0	98	STAT2	98	0	2
HNRNPUL1	100	0	0	TAF10	100	0	0
IRF1	100	0	0	TAF3	100	0	0
KAT2B	98	0	2	TAF5	100	0	0
KDM1A	100	0	0	THRA	61	15	24
KEAP1	100	0	0	VDR	20	5	75
KHDRBS1	100	0	0	VEGFA	100	0	0
KHDRBS3	0	0	100	YY1	100	0	0
KLF4	34	2	64	ZFPM1	33	13	54
LCOR	100	0	0				

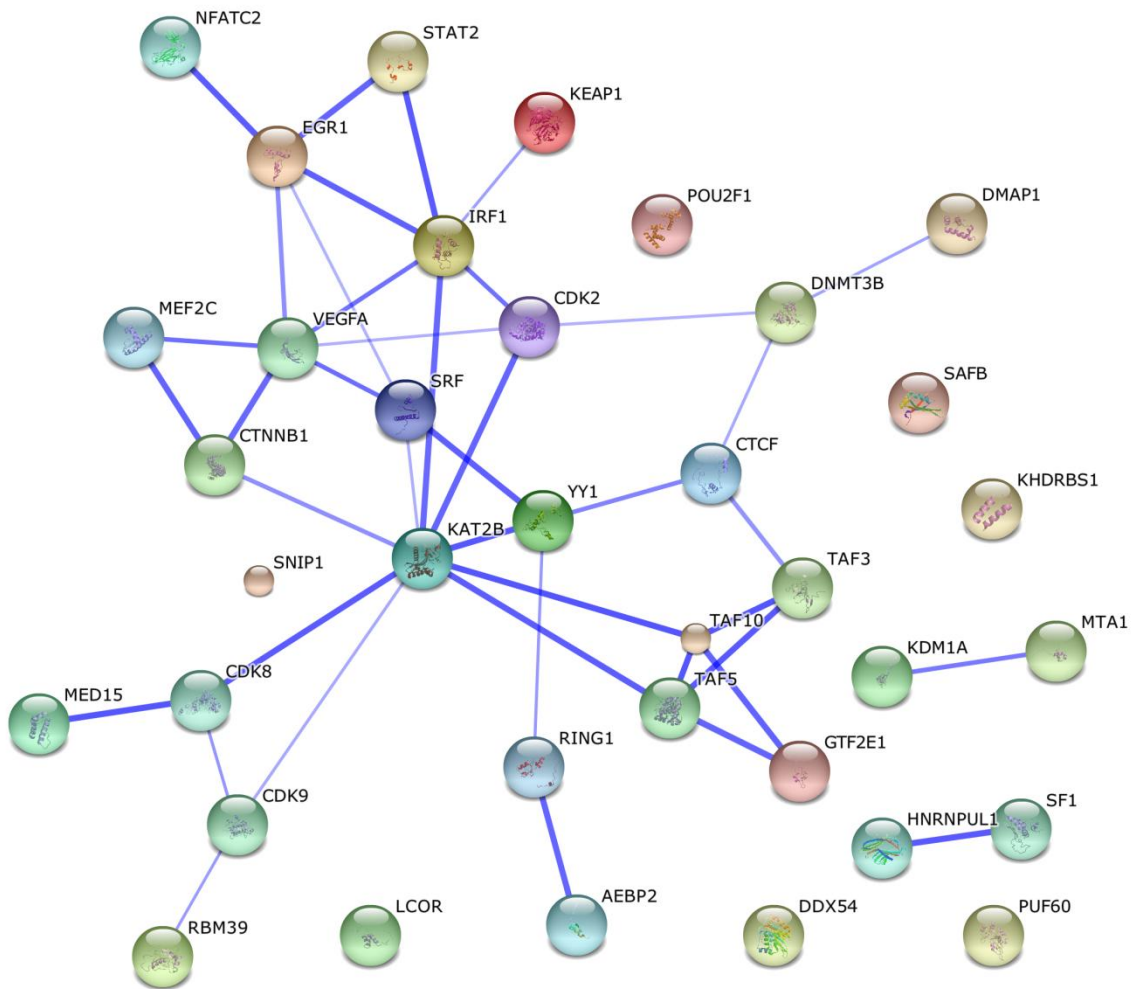
M – marginalno prisutna ekspresija; O – odsutna ekspresija; P – prisutna ekspresija.

4.1.2.5. Analiza koekspresije integrativnog algoritma primenjena na WT1

Da bi neka PPI mogla da se odigra u realnim okolnostima u ćeliji, oba proteina interaktora moraju biti eksprimirana u toj ćeliji. Shodno tome, proteini kandidati za nove WT1 interaktore se dalje filtriraju na osnovu koekspresije u ćelijama u kojima se planiraju biološka ispitivanja. Model sistem za ispitivanje efekata WT1 u mijeloidnim malignitetima je ćelijska linija K562, sa BCR-ABL pozitivnim ćelijama CML tipa. Za analizu ekspresije WT1 i potencijalnih novih interaktora u K562, korišćeni su *microarray* podaci uskladišteni u GEO bazi. Analiza prisustva ekspresije gena u uzorku daje jedan od tri moguća rezultata: prisutna ekspresija (P), marginalno prisutna ekspresija (M) i odsutna ekspresija (O).

WT1 je eksprimiran u svim analiziranim uzorcima K562 ćelijske linije. Rezultati analize koekspresije 51 proteina iz skupa LISTA4 sa WT1 u K562 ćelijskoj liniji su prikazani u Tabeli 8. Kandidatima za nove WT1 interaktore se mogu smatrati oni proteini koji su eksprimirani (P+M) u najvećem broju analiziranih uzoraka, a odabrana je granica od 80%. Ovaj uslov ispunjava 36 proteina ([Slika 10](#)) i oni predstavljaju najbolje kandidate za nove interaktore WT1, identifikovane integrativnim algoritmom.

Naknadnim eksperimentalnim testiranjem, koimunoprecipitacijom, Roberts i saradnici su potvrdili fizičku interakciju između WT1 i CDK9 u K562 ćelijama (Roberts i sar. rad u pripremi).

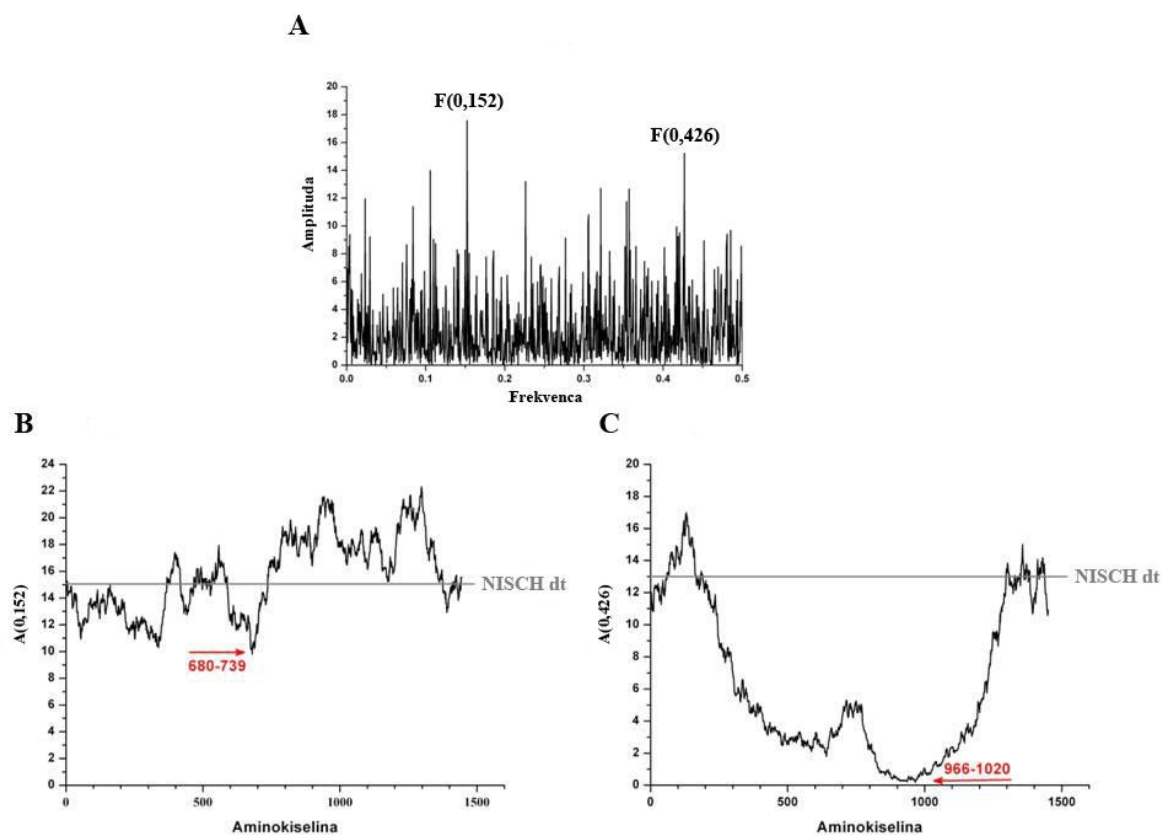


Slika 10. Predvideni novi interaktori WT1, koekspimirani sa WT1 u K562 ćelijskoj liniji, i njihove međusobne veze; debljina linija zavisi od STRING skora poverenja.

4.2. Predviđanje i karakterizacija funkcionalnih domena proteina NISCH

4.2.1. Predviđanje funkcionalnih domena proteina NISCH

NISCH predstavlja novi target za ciljanu terapiju mijeloidnih maligniteta. Identifikacija domena koji učestvuju u terapijskom, apoptotskom, delovanju NISCH će ubrzati generisanje racionalno dizajniranih terapeutika. U prvom koraku identifikacije funkcionalnih domena NISCH, generisan je informacioni spektar ovog proteina i identifikovane su dve karakteristične frekvence, $F(0,152)$ i $F(0,426)$ ([Slika 11A](#)). Da bi se identifikovali delovi sekvence koji najviše doprinose intenzitetu ovih dominantnih spektralnih karakteristika, izvršena je serija *in silico* delecija pomoći programa *AKSkener*. N uzastopnih aminokiselina se briše iz proteinske sekvence, pa se iterativno generiše skup novih sekvenci tako što za poziciju od koje počinje isecanje uzima redom sve pozicije u ulaznoj sekvenci od prve do poslednje moguće. *AKSkener* za svaku novodobijenu sekvencu računa informacioni spektar i vrednosti amplitude i S/N na zadatoj frekvenci. Na osnovu toga se može zaključiti koji region proteina najviše doprinosi intenzitetu pika na nekoj frekvenci jer njegovo isecanje dovodi do najvećeg smanjenja pika. Ovako su identifikovani domeni čije ukljanjanje izaziva najznačajniji pad amplitude na karakterističnim frekvencama u spektru protina NISCH. Korišćena veličina regiona za isecanje je $N = 55-60$ aminokiselina. Domen između 680 i 739 pozicije u NISCH sekvenci najviše doprinosi intenzitetu karakteristične frekvence $F(0,152)$ ([Slika 11B](#)). Domen 966-1020 je odgovoran za amplitudu na frekvenci $F(0,426)$, pri čemu delecija ovog domena dovodi do potpunog pada amplitude na ovoj frekvenci, praktično na nulu, ukazujući na važnost ovog domena za funkciju NISCH-a ([Slika 11C](#)).



Slika 11. Spektralna analiza proteina NISCH. A. Spektar proteina NISCH, sa označenim karakterističnim frekvencama, F(0,152) i F(0,426). B. Domen koji najviše doprinosi intenzitetu amplitude na frekvenci F(0,152). C. Domen koji najviše doprinosi intenzitetu amplitude na frekvenci F(0,426).

Naknadna 3D-QSAR analiza je potvrdila da se u okviru ovde identifikovanih domena NISCH-a nalaze potencijalna mesta vezivanja I₁-IR liganada (Nikolic i sar. 2013).

4.2.2. Funkcionalna karakterizacija domena proteina NISCH

Osnovni princip u predviđanju funkcija proteina na osnovu njegove sekvence je zasnovan na identifikaciji sličnih proteina (po sekvenci, strukturi ili nekim drugim

parametrima) sa poznatom funkcijom (Whisstock i Lesk 2003). Za funkcionalnu karakterizaciju domena proteina NISCH koristi se detekcija proteina sa poznatim funkcijama, koji imaju F(0,152) i F(0,426) kao dominantne frekvence u svom spektru (slična spektralna svojstva).

Prvo je pretražena UniProt baza humanih proteina pomoću programa *FilterDFTFFTBase*. Tako je identifikovano 553 proteina koji imaju F(0,152) kao jednu od dve dominantne frekvence u spektru (Prilog 3). Da bi se utvrdilo koji su biološki procesi u koje su uključeni proteini u ovom skupu, izvršena je *enrichment* analiza pomoću DAVID platforme. Korišćen je alat *Gene_Ontology Enrichment*, sa terminima GO baze koji se odnose na biološke procese. Deset najvažnijih bioloških procesa je prikazano u Tabeli 9. S obzirom da je domen 680-739 proteina NISCH asociran sa F(0,152), sledi da je taj domen ključan za ulogu NISCH u biološkim procesima u Tabeli 9.

Tabela 9. GO termini koji reprezentuju biološke procese u koje je uključen domen 680-739 proteina NISCH (asociran sa F(0,152)); *enrichment* na svim terminima je statistički značajno ($p \leq 0,01$).

GO:0031400~negative regulation of protein modification process
GO:0030433~ER-associated protein catabolic process
GO:0031397~negative regulation of protein ubiquitination
GO:0031396~regulation of protein ubiquitination
GO:0031399~regulation of protein modification process
GO:0042062~long-term strengthening of neuromuscular junction
GO:0008582~regulation of synaptic growth at neuromuscular junction
GO:0051963~regulation of synaptogenesis
GO:0050807~regulation of synapse organization
GO:0050803~regulation of synapse structure and activity

Da bi se dobio pregled širih kategorija u koje spadaju dobijeni GO termini, GO termini iz Tabele 9 su tretirani kao *GO Slim* i izvršena je vizualizacija dela celokupne GO ontologije oko ovih termina, pomoću *QickGO* programa ([Slika 12](#)). Dobijeni GO termini, asocirani sa F(0,152), se mogu podeliti u 2 grupe: i) procesi uključeni u posttranslacione

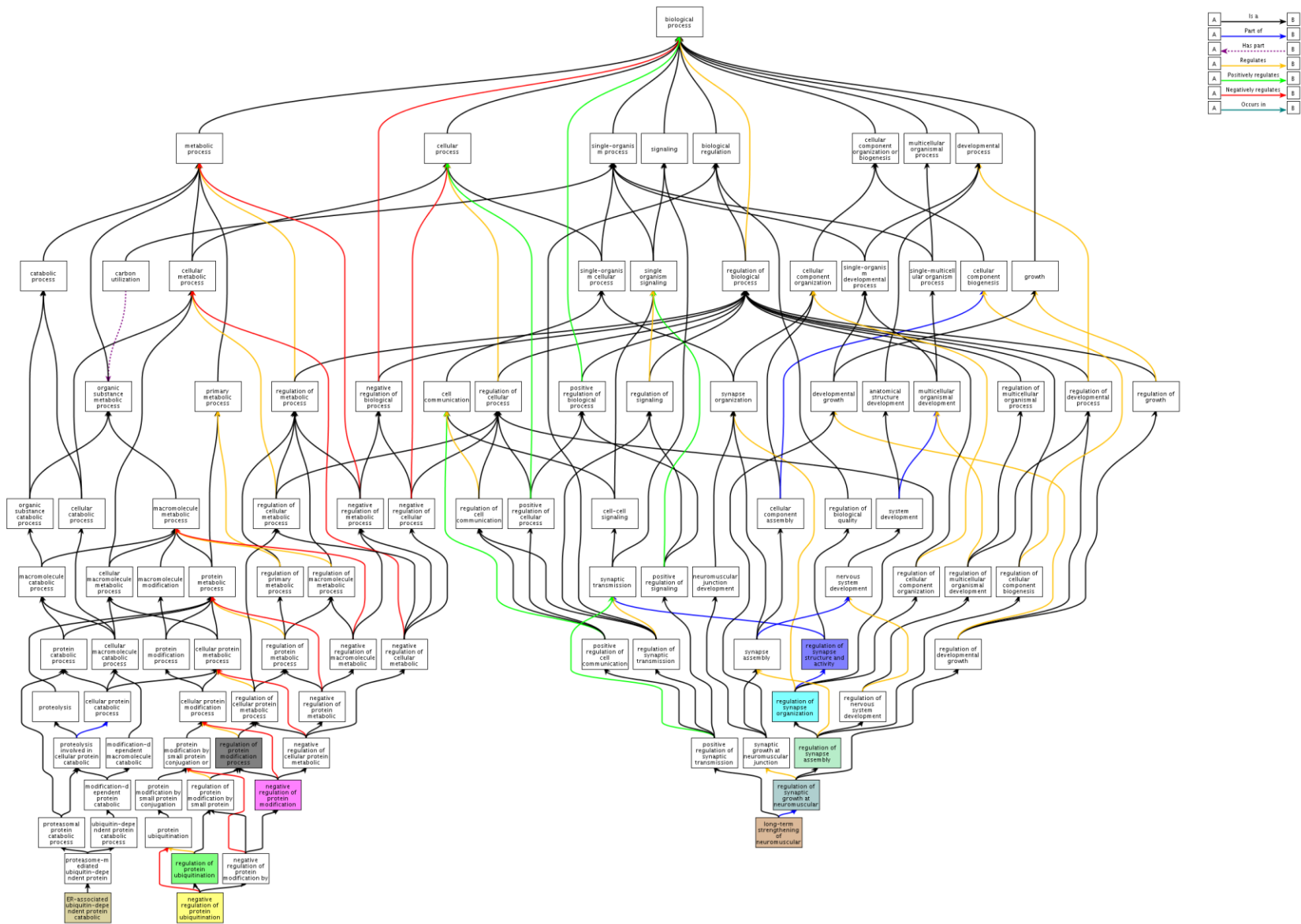
modifikacije proteina, specifično ubikvitinaciju, i ii) procesi uključeni u razvoj nervnih ćelija, tačnije stvaranje sinapsi.

Iste analize u *FilterDFTFFTBaze* programu su urađene i za frekvencu F(0,426). Tako je identifikovano 820 proteina koji imaju F(0,426) kao jednu od dve dominantne frekvence u spektru (Prilog 4). *Gene_Ontology Enrichment* alat DAVID platforme je identifikovao osam najvažnijih bioloških procesa u koje su uključeni ovi proteini (Tabela 10). S obzirom da je domen 966-1020 proteina NISCH asociran sa F(0,426), sledi da je on ključan za ulogu NISCH u biološkim procesima navedenim u Tabeli 10.

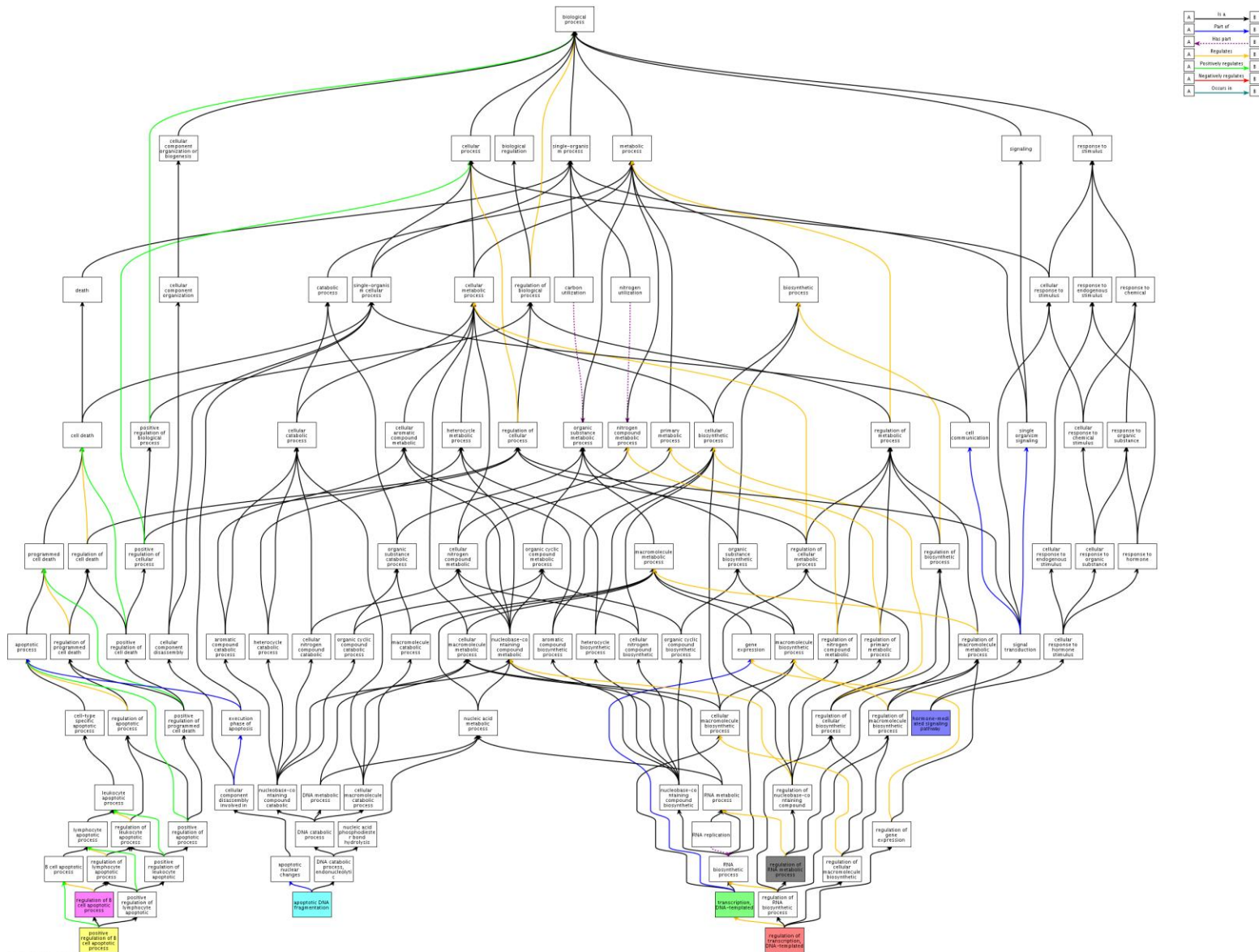
Tabela 10. GO termini koji reprezentuju biološke procese u koje je uključen domen 966-1020 proteina NISCH (asociran sa F(0,426)); *enrichment* na svim terminima je statistički značajno ($p \leq 0,01$).

GO:0006350~transcription
GO:0006355~regulation of transcription, DNA-dependent
GO:0045449~regulation of transcription
GO:0051252~regulation of RNA metabolic process
GO:0002902~regulation of B cell apoptosis
GO:0006309~DNA fragmentation involved in apoptosis
GO:0002904~positive regulation of B cell apoptosis
GO:0009755~hormone-mediated signaling

Vizualizacija ovih termina pomoću *QickGO* programa je prikazana na [Slici 13](#). Identifikovane su 3 kategorije bioloških procesa asociranih sa F(0,426): i) regulacija transkripcije, ii) apoptoza, koja obuhvata dva specifična procesa – fragmentaciju DNK uključenu u apoptozu i regulaciju apoptoze B ćelija, i iii) prenos signala, specifično hormonima-posredovan prenos signala.



Slika 12. Vizualizacija GO termina, koji reprezentuju biološke procese u koje je uključen domen 680-739 proteina NISCH (asociran sa F(0,152)).



Slika 13. Vizualizacija GO termina, koji reprezentuju biološke procese u koje je uključen domen 966-1020 proteina NISCH (asociran sa F(0,426)).

4.3. Predviđanje funkcionalnih efekata AAS u nCFD regionima epigenetičkih regulatora ASXL1, EZH2, DNMT3A i TET2

4.3.1. Lista AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2

Za pravljenje kolekcije AAS u ASXL1, EZH2, DNMT3 i TET2 je korišćena literatura, COSMIC i dbSNP baze podataka. Preuzeti su podaci i o tipu mijeloidnog maligniteta sa kojim su asocirane štetne mutacije, literaturi i patogenom statusu AAS. Patogeni status je smatran mutantnim ukoliko je pokazano da je AAS stečena u toku razvoja bolesti (somatska mutacija), tj. ukoliko je prisutna u obolelim ćelijama, ali ne i u drugim tkivima istog pacijenta (pr. koža, limfociti, itd.). Za AAS je označeno da ima neutralan patogeni status (SNP) ukoliko je pokazano njegovo prisustvo i u obolelom i u nekom drugom tkivu pacijenta, ili ukoliko je utvrđena učestalost pojavljivanja tog polimorfizma u zdravoj populaciji.

Lista svih AAS u ASXL1, EZH2, DNMT3 i TET2 je data u Prilogu 5 i rezimirana u Tabeli 11. Ovaj skup je označen kao SVE_AAS i sadrži 314 supstitucija – 120 SNP-ova i 194 mutacija. Najveći broj mutacija je detektovan kod AML pacijenata, oko 45%. Kod pacijenata sa MDS, MPN i MDS/MPN detektovano je 12%, 13% i 7% mutacija, redom. Ostale mutacije su detektovane u dva ili više mijeloidna maligniteta. Takođe, više od polovine AAS su TET2 – 62% mutacija i 38% SNP-ova, ~24% su ASXL1, 8% su EZH2 i 15% su DNMT3A AAS.

Tabela 11. Pregled SVE_AAS, skupa svih aminokiselinskih supstitucija u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2.

	SNP (n=120)		MUT (n=194)	
	nCFD	CFD	nCFD	CFD
ASXL1 (n=76)	59	4	12	1
EZH2 (n=25)	4	2	6	13
DNMT3A (n=47)	3	3	6	35
TET2 (n=166)	42	3	27	94
Total (n=314)	108	12	51	143

CFD – *Conserved Functional Domains*; MUT – mutacija; nCFD – *not CFD*; SNP – *Single Nucleotide Polymorphism*.

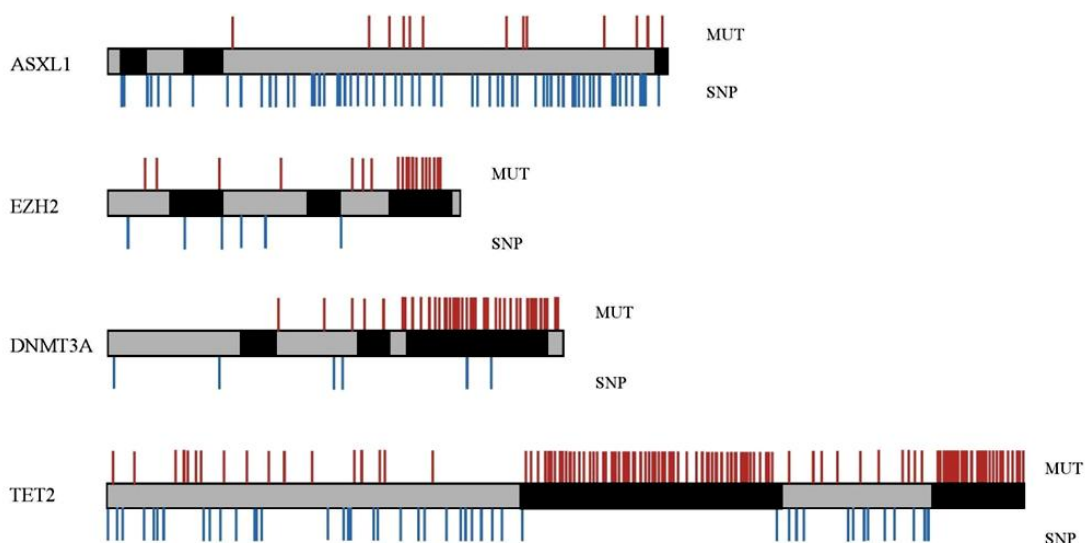
S obzirom da su predmet istraživanja AAS koje su pozicionirane u nCFD regionima epigenetičkih regulatora ASXL1, EZH2, DNMT3A i TET2, iz SVE_AAS skupa je izdvojen podskup sa ovim supstitucijama – nCFD_AAS. Pozicija CFD i nCFD je preuzeta iz literature (Tabela 12). Ovaj podskup je obuhvatio 51 od 194 mutacije (26%) u epigenetičkim regulatorima, što je manje od očekivanog na osnovu literaturnih podataka zasnovanih na svim proteinima asociranim sa kancerima (Yue i sar. 2010; Peterson i sar. 2012). S druge strane, očekivano, oko 90% svih SNP-ova u ovim epigenetičkim regulatorima su pozicionirani u nCFD.

Kao što se vidi iz Tabele 12, nCFD_AAS sadrži 51 mutaciju i 108 SNP-ova. Oko 45% su ASXL1 supstitucije – 55% SNP-ova i 24% mutacija. ~ 6% AAS su EZH2, 6% DNMT3A i 43% TET2. Distribucija polimorfizama duž sekvenci ovih proteina je prikazana na [Slici 14](#).

Tabela 12. Epigenetički regulatori ASXL1, EZH2, DNMT3A i TET2: njihovi ID u UniProt bazi, konzervirani funkcionalni domeni i odgovarajuća literatura.

Protein	UniProt ID	CFD	Pozicija (ak)	Literatura
ASXL1	Q8IXJ9	HARE ASXH PHD	11-83 241-369 1506-1541	Gelsi-Boyer i sar. 2012
EZH2	Q15910	SANT1 SANT2 SET	159-250 433-481 617-738	Jankowska i sar. 2011
DNMT3A	Q9Y6K1	PWWP PHD MTase	290-348 536-589 638-908	Jankowska i sar. 2011
TET2	Q6N021	BOX1 BOX2	1104-1478 1845-2002	Langemeijer i sar. 2009

CFD – conserved functional domain.



Slika 14. Distribucija mutacija u mijeloidnim malignitetima (crveno) i SNP-ova (plavo) duž sekvenci ASXL1, EZH2, DNMT3A i TET2 epigenetičkih regulatora; crnim su označeni konzervirani funkcionalni domeni.

4.3.2. Efikasnost PolyPhen-2 i SIFT u predviđanju funkcionalnih efekata nCFD AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2

PolyPhen-2 i SIFT su najšire korišćeni MSA-zasnovani alati za predikciju funkcionalnih efekata AAS. Oni su upotrebljeni za predviđanje funkcionalnih efekata AAS pozicioniranih u nCFD regionima proteina. U ovom delu istraživanja se ispituje njihova efikasnost u predikciji funkcionalnih efekata AAS pozicioniranih u nCFD u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2. Performansa je iskazana na nCFD_AAS pomoću tri statistička parametra: tačnost, senzitivnost i specifičnost.

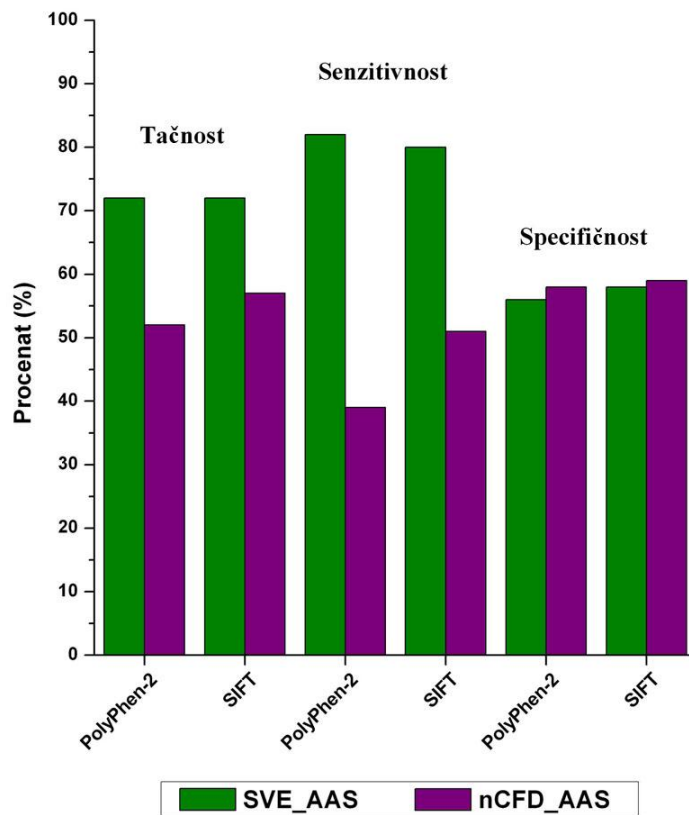
Kao što se vidi iz Tabele 13, PolyPhen-2 i SIFT nisu efikasni za predviđanje funkcionalnih efekata nCFD AAS u epigenetičkim regulatorima (rezultati Fišerovog testa pokazuju da je $p > 0,05$). Tačnost, senzitivnost i specifičnost oba alata su niže od 60%. Oba alata su imala tendenciju da funkcionalni efekat nCFD AAS proglašavaju neutralnim: od 159 AAS pozicioniranih u nCFD, PolyPhen-2 je 94 (oko 59%) klasifikovao kao SNP-ove, a SIFT 89 (oko 56%). Ova tendencija ima posebno negativan uticaj na senzitivnost. PolyPhen-2 je za 61% AAS čiji je realan patogeni status mutantni, predvideo neutralan funkcionalan efekat. A SIFT rezultati obuhvataju 49% lažno negativnih rezultata.

Tabela 13. Statistički parametri efikasnosti predviđanja funkcionalnih efekata aminokiselinskih supstitucija PolyPhen-2 i SIFT na nCFD_AAS skupu (p vrednosti su dobijene Fišerovim testom).

	TP	TN	LP	LN	Senzitivnost	Specifičnost	Tačnost
PolyPhen-2 ($p=0,863$)	20	63	45	31	0,39	0,58	0,52
SIFT ($p=0,236$)	26	64	44	25	0,51	0,59	0,57

LN – lažno negativne predikcije; LP – lažno pozitivne predikcije; TN – tačno negativne predikcije; TP – tačno pozitivne predikcije.

Niska efikasnost PolyPhen-2 i SIFT u predviđanju funkcionalnih efekata substitucija u nCFD_AAS skupu može biti posledica: i) niske efikasnosti ovih alata u predviđanju funkcionalnih efekata AAS u epigenetičkim regulatorima ili ii) niske efikasnosti u predviđanju funkcionalnih efekata nCFD pozicioniranih AAS, *per se*. S obzirom da PolyPhen-2 i SIFT nikada nisu testirani specifično na funkcionalnom podskupu, epigenetičkim regulatorima, testirali smo prvu hipotezu korišćenjem SVE_AAS skupa podataka. Pokazalo se da je senzitivnost, pa posledično i tačnost, ovih alata značajno veća u odnosu na nCFD_AAS ([Slika 15](#)). Ovo ukazuje na to da se može odbaciti prva i prihvatiti druga pretpostavka, tj. da predikcije PolyPhen-2 i SIFT u velikoj meri zavise od konzerviranosti pozicije AAS u sekvenci proteina, bar u slučaju epigenetičkih regulatora, tj. od lokalizacije AAS u CFD ili nCFD. To je potvrđeno Spirmanovom korelacionom analizom. Ona je pokazala značajne korelacije velikog intenziteta ($r \geq 0,50$) CFD/nCFD pozicije AAS i PolyPhen-2 predikcija i izuzetno značajne korelacije srednjeg intenziteta ($0,30 \leq r \leq 0,49$) sa SIFT predikcijama (Tabela 14).



Slika 15. Poređenje tačnosti, senzitivnosti i specifičnosti alata PolyPhen-2 i SIFT u predviđanju funkcionalnih efekata aminokiselinskih supstitucija iz SVE_AAS (zeleni stubići) i nCFD_AAS (ljubičasti stubići) skupova.

Tabela 14. Spirmanove korelacije CFD/nCFD pozicije aminokiselinskih supstitucija i PolyPhen-2 i SIFT predikcija; sve korelacije su statistički značajne ($p < 0,001$).

	SIFT	PolyPhen-2
CFD/nCFD	0,45	0,58
PolyPhen-2	0,48	

CFD – conserved functional domain; nCFD – not CFD.

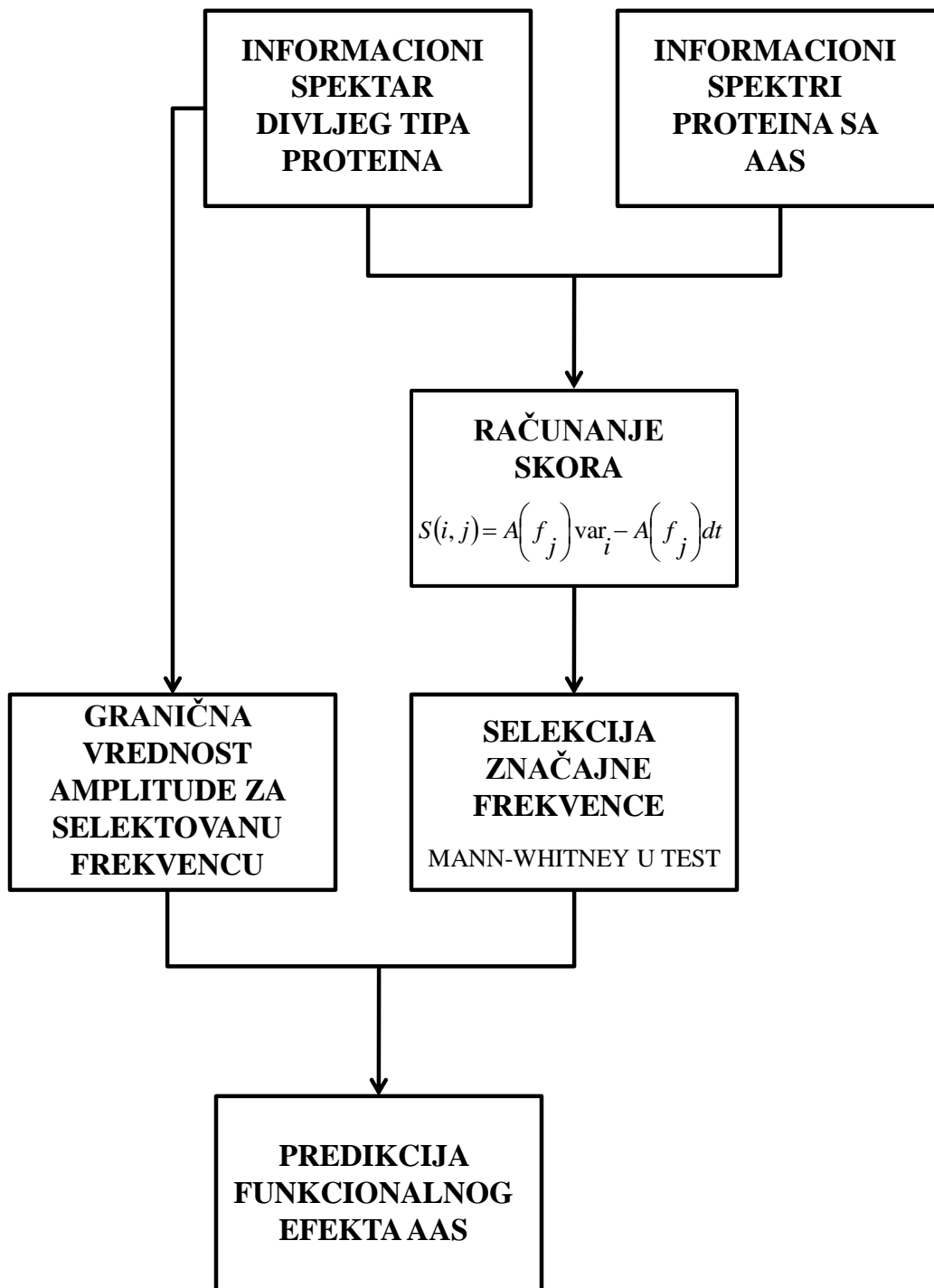
Ovi rezultati pokazuju da MSA-zasnovani alati nisu efikasni u predviđanju funkcionalnih efekata nCFD pozicioniranih AAS u epigenetičkim regulatorima i da postoji potreba za drugačijim pristupom.

4.3.3. ISM algoritam za predikciju funkcionalnih efekata AAS

ISM algoritam (Gemovic i sar. 2013) za predikciju funkcionalnih efekata AAS je prikazan na [Slici 16](#) i podrazumeva pet koraka:

- i) Generisanje informacionih spektara
- ii) Računanje ISM skora
- iii) Selekcija značajne frekvence
- iv) Određivanje granične vrednosti amplitude za selektovanu frekvencu
- v) Predikcija funkcionalnog efekta AAS

Pretpostavke koje leže u osnovi ovog algoritma su: i) efekat AAS na karakteristike proteinskog spektra nije u korelaciji sa evolucionom konzerviranošću AAS pozicije i ii) efekat AAS na proteinsku funkcije se može predvideti pomoću ISM.



Slika 16. ISM algoritam za predikciju funkcionalnih efekata aminokiselinskih supstitucija.

4.3.3.1. Generisanje informacionih spektara

Efekat AAS na karakteristike spektra nekog proteina se ispituje poređenjem sa spektralnim svojstvima divljeg tipa (dt) tog proteina. Stoga se, u prvom koraku, kreira informacioni spektar dt proteina, pomoću programa ISM platforme – *ProteinSpektar*. Korišćen je DFT metod za transformaciju EIIP kodiranih proteinskih sekvenci u informacione spektre.

4.3.3.2. Računanje ISM skora

Za svaku sekvencu sa AAS se računa ISM skor na svim frekvencama koje su prisutne u informacionom spektru dt proteina. Skor se računa kao odstupanje vrednosti amplitude na određenoj frekvenci u informacionom spektru sekvence sa AAS u odnosu na dt:

$$S(i, j) = A(f_j)var_i - A(f_j)dt, \quad i = 1, 2, \dots, N \quad j = 1, 2, \dots, M \quad (9)$$

gde je N broj ispitivanih AAS u tom proteinu, a M je broj frekvenci u informacionom spektru dt. $A(f_j)var_i$ vrednosti se računaju pomoću programa ISMplatforme – *MutacijeDFT*.

4.3.3.3. Selekcija značajne frekvence

Da bi se odredila statistička značajnost neke frekvence za razlikovanje sekvenci sa SNP-ovima i onih sa mutacijama, koristi se neparametrijski Mann-Whitney U test. Statistička analiza se radi na N sekvenci sa AAS. Prvo se analizira frekvencia koja ima maksimalnu amplitudu u informacionom spektru dt proteina. Ukoliko ona nije statistički značajna, ista analiza se sprovodi na drugim frekvencama redom po kojim im opada amplituda u informacionom spektru dt proteina. Selektuje se prva frekvencia za koju Mann-Whitney U test pokaže statističku značajnost. Da bi se proverilo da li promene amplitude na selektovanoj frekvenci zaista ne zavise od toga da li je AAS pozicionirana u CFD ili nCFD, radi se virtuelno Ala skeniranje, tj. zamena svake pojedinačne aminokiseline alaninom. Za to se koristi program ISM platforme – *AKSkener*.

4.3.3.4. Određivanje granične vrednosti amplitude za selektovanu frekvencu

Informacioni spektar dt proteina predstavlja referentni okvir, a funkcionalno značajne mutacije menjaju vrednosti amplituda na svim frekvencama u spektru. Kao granična vrednost uzima se intenzitet amplitude na selektovanoj frekvenci u spektru dt proteina. Promene vrednosti amplitude, do kojih dovode štetne mutacije, mogu ići u pravcu povećanja ili smanjenja, ali je bitno naznačiti da neutralni SNP-ovi ne mogu imati isti pravac promena kao ove funkcionalno značajne mutacije.

4.3.3.5. Predikcija funkcionalnog efekta AAS

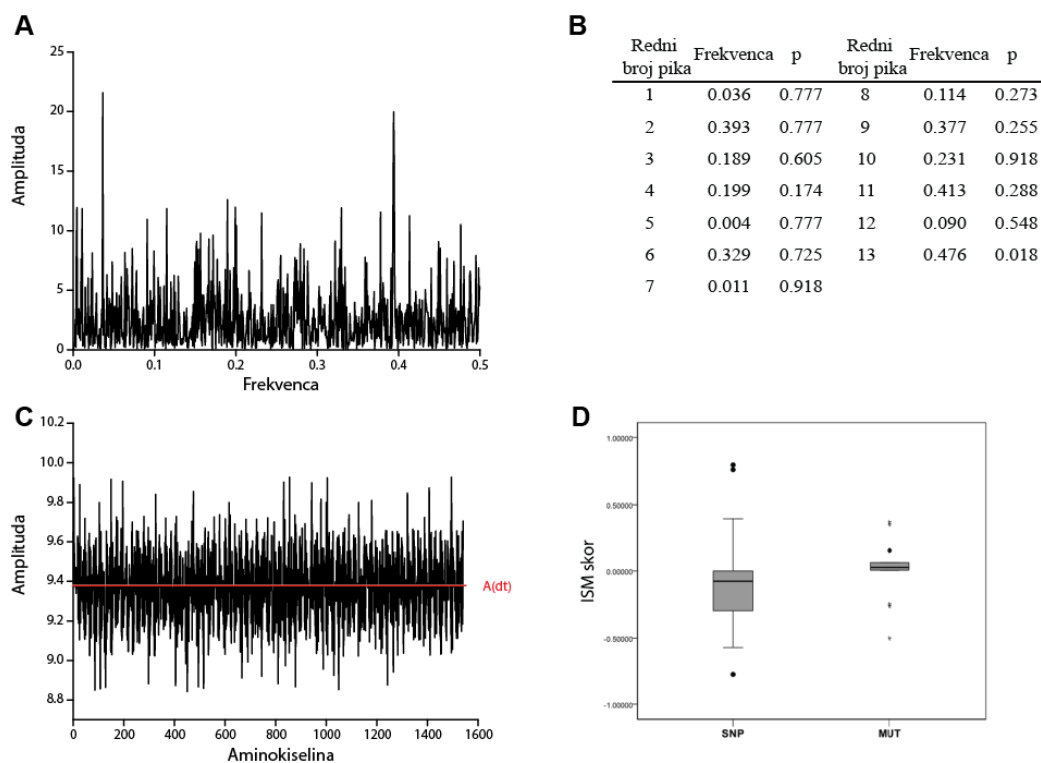
Utvrđuje se da li najveći broj štetnih mutacija povećava ili smanjuje amplitudu na selektovanoj frekvenci u odnosu na graničnu vrednost. Predikcija se vrši u skladu s tim, na osnovu vrednosti ISM skora. Npr. ukoliko je utvrđeno da najveći broj štetnih mutacija povećava vrednost amplitude na selektovanoj frekvenci u odnosu na graničnu vrednost, predviđanje za sve AAS sa pozitivnim ISM skorom će glasiti da su štetne mutacije, a za AAS sa ISM skorom manjim ili jednakim graničnoj vrednosti, predviđanje će glasiti da su neutralni SNP-ovi.

4.3.4. Predviđanje funkcionalnih efekata AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2 pomoću ISM algoritma

4.3.4.1. Predikcija funkcionalnih efekata AAS u proteinu ASXL1

Informacioni spektar ASXL1 proteina je konstruisan pomoću programa *ProteinSpektar* ([Slika 17A](#)). Zatim su, pomoću programa *MutacijeDFT*, za svih 76 sekevenci sa AAS izračunati ISM skorovi na svakoj frekvenci iz informacionog spektra dt proteina ASXL1. U trećem koraku su skorovi na frekvenci sa najvišom amplitudom, F(0,036), testirani Mann-Whitney U testom. S obzirom da ova frekvenca nije značajno razlikovala sekvence sa mutacijama i SNP-ovima, isti statistički test je primenjen na sledeću frekvencu po veličini

amplitude u spektru ASXL1. Ova procedura je ponavljana dok nije identifikovana frekvencija $F(0,476)$, na kojoj ISM skorovi značajno razlikuju sekvence sa mutacijama u ASXL1 od sekvenci sa SNP-ovima ($p=0,018$) ([Slika 17B](#)). Pre finalnog koraka predikcije funkcionalnih efekata AAS u nCFD proteina ASXL1, potvrđeno je, pomoću virtuelnog Ala skeniranja *AKSkener* programom, da efekat na vrednost amplitude na $F(0,476)$ ne zavisi ($p = 0,884$) od pozicije AAS u CFD/nCFD ([Slika 17C](#)). U poslednjem koraku, pokazalo se da 77% sekvenci sa mutacijama ima više vrednosti amplitude na ovoj frekvenci u odnosu na dt. Niže vrednosti amplitude na $F(0,476)$ u odnosu na dt proteina ASXL1 ima oko 75% sekvenci sa SNP-ovima ([Slika 17D](#)).

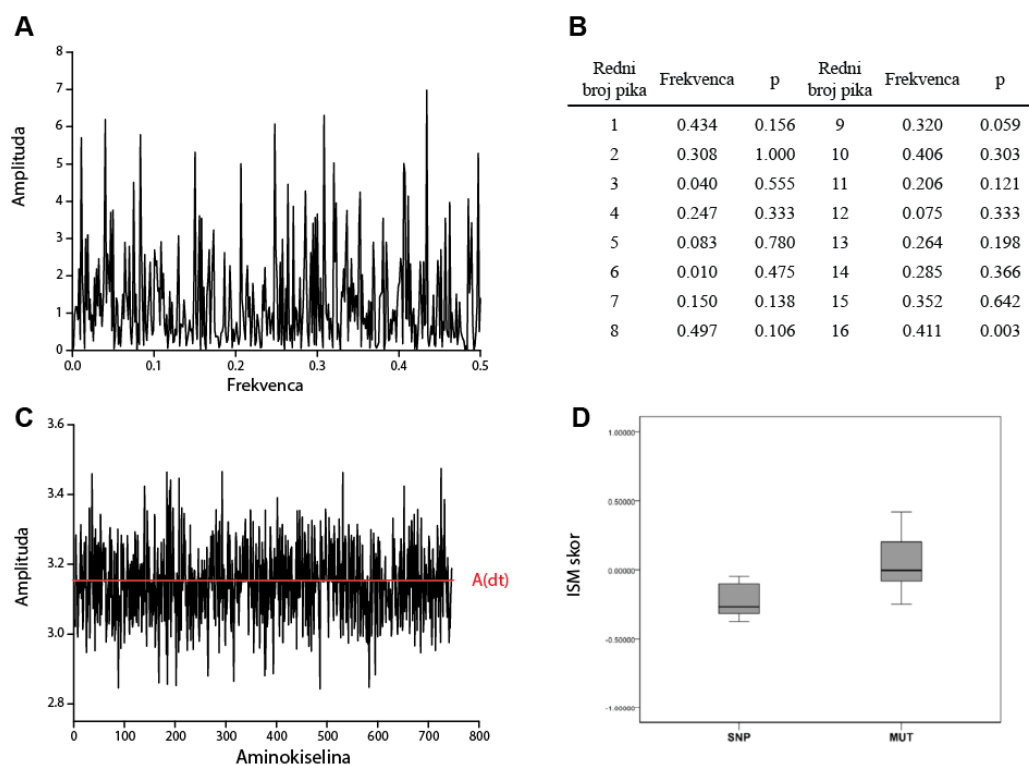


Slika 17. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu ASXL1.

A. Informacioni spektar divljeg tipa proteina ASXL1. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala skeniranje na frekvenci $F(0,476)$. D. Raspodela ISM skorova za SNP-ove i mutacije u ASXL1.

4.3.4.2. Predikcija funkcionalnih efekata AAS u proteinu EZH2

Primenom analize, koja je opisana u prethodnom paragrafu, na sekvencu EZH2 ([Slika 18A](#)) i njene varijacije, identifikovana je frekvencija F(0,411), na kojoj ISM skorovi značajno razlikuju sekvence sa mutacijama od sekvenci sa SNP-ovima ($p=0,003$) ([Slika 18B](#)). Pomoću virtuelnog Ala skeniranja *AKSkener* programom, potvrđeno je da efekat na vrednost amplitude na F(0,411) ne zavisi ($p = 1,000$) od pozicije AAS u CFD/nCFD ([Slika 18C](#)). U poslednjem koraku, pokazalo se da svih šest sekvenci sa SNP-ovima ima nižu vrednost amplitude na ovoj frekvenci u odnosu na divlji tip, dok približno polovina sekvenci sa mutacijama ima ove vrednosti više od divljeg tipa EZH2 proteina ([Slika 18D](#)).

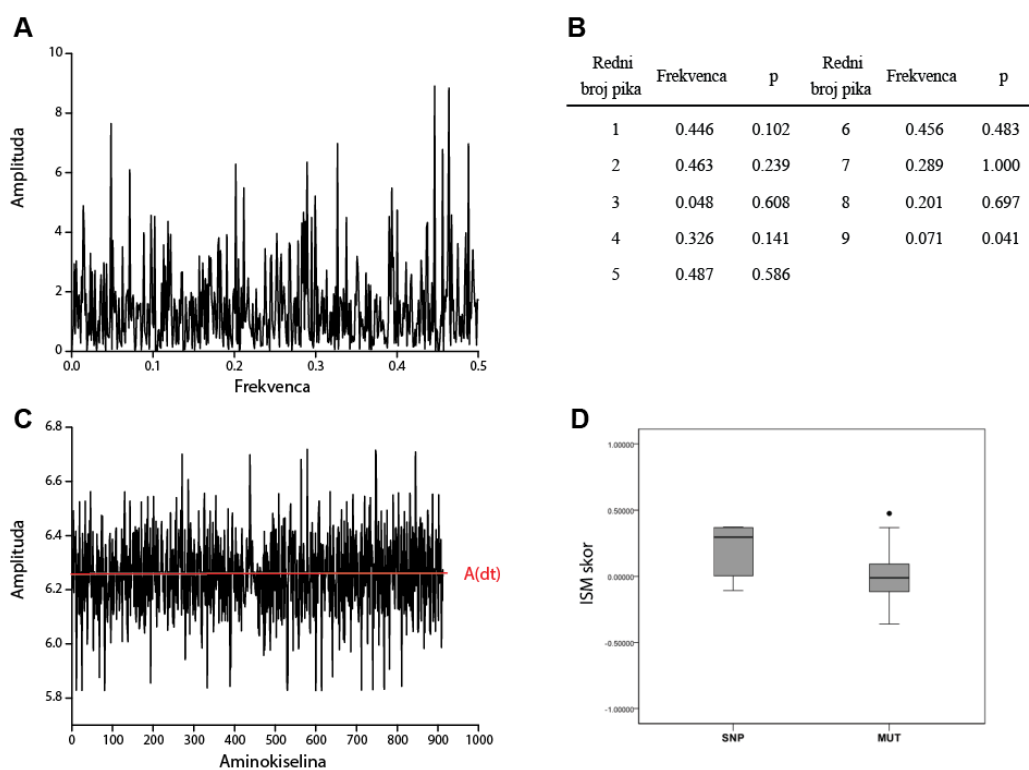


Slika 18. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu EZH2.

A. Informacioni spektar divljeg tipa proteina EZH2. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala skeniranje na frekvenci F(0,411). D. Raspodela ISM skorova za SNP-ove i mutacije u EZH2.

4.3.4.3. Predikcija funkcionalnih efekata AAS u proteinu DNMT3A

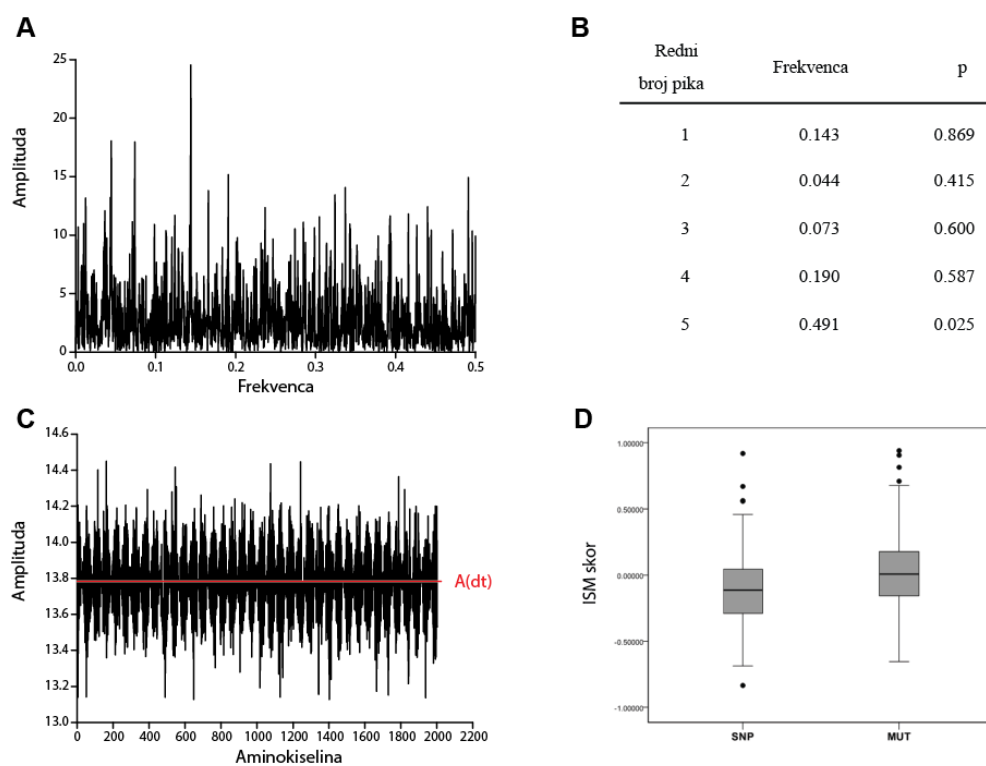
ISM analiza je pokazala da je za razlikovanje sekvenci sa mutacijama od onih sa SNP-ovima u DNMT3A značajna frekvencija $F(0,071)$ ($p=0,041$) (Slike 19A i B). Pomoću virtuelnog Ala skeniranja *AKSkener* programom, potvrđeno je da efekat na vrednost amplitude na $F(0,071)$ ne zavisi ($p = 0,164$) od pozicije AAS u CFD/nCFD (Slika 19C). Za razliku od mutacija u ASXL1 i EZH2 proteinima, koje su povećavale vrednost amplitude na značajnoj frekvenci u odnosu na divlji tip, mutacije u DNMT3A je smanjuju. Nešto više od polovine sekvenci sa mutacijama (51%) ima niže amplitude na frekvenci $F(0,071)$ u odnosu na dt, dok većina sekvenci DNMT3A sa SNP-ovima (83%) ima povećane vrednosti amplitude na ovoj frekvenci u odnosu na dt (Slika 19D).



Slika 19. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu DNMT3A. A. Informacioni spektar divljeg tipa proteina DNMT3A. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala skeniranje na frekvenci $F(0,071)$. D. Raspodela ISM skorova za SNP-ove i mutacije u DNMT3A.

4.3.4.4. Predikcija funkcionalnih efekata AAS u proteinu TET2

Informacioni spektar dt proteina TET2, konstruisan pomoću programa *ProteinSpektar*, je prikazan na [Slici 20A](#). Pomoću programa *MutacijeDFT*, za svih 161 sekvencu sa AAS izračunati su ISM skorovi na svakoj frekvenci iz informacionog spektra dt proteina TET2. Mann-Whitney U test je identifikovao frekvencu F(0,491) kao značajnu za razlikovanje sekvenci sa mutacijama u TET2 od sekvenci sa SNP-ovima ($p=0,025$) ([Slika 20B](#)). Pomoću virtuelnog Ala skeniranja *AKSkener* programom, potvrđeno je da efekat na vrednost amplitude na F(0,491) ne zavisi ($p = 0,675$) od pozicije AAS u CFD/nCFD ([Slika 20C](#)). Oko 60% TET2 sekvenci sa SNP-ovima ima nižu vrednost amplitude, a 55% sekvenci sa mutacijama ima višu vrednost amplitude na ovoj frekvenci u odnosu na dt ([Slika 20D](#)).



Slika 20. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu TET2.

A. Informacioni spektar divljeg tipa proteina TET2. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala skeniranje na frekvenci F(0,491). D. Raspodela ISM skorova za SNP-ove i mutacije u TET2.

4.3.5. Poređenje efikasnosti ISM algoritma u predviđanju funkcionalnih efekata nCFD AAS u epigenetičkim regulatorima sa PolyPhen-2 i SIFT

Performansa ISM algoritma u predikciji funkcionalnih efekata nCFD AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2 je procenjena pomoću parametara: tačnost, senzitivnost i specifičnost (Tabela 14). Korišćen je skup supstitucija nCFD_AAS. Ukupna tačnost ISM algoritma je veća za 17% i 12% u odnosu na PolyPhen-2 i SIFT, redom (v. Tabele 13 i 15). Specifičnost ISM algoritma je veća za 12% i 11%, dok je senzitivnost veća za 26% i 14% u odnosu na PolyPhen-2 i SIFT, redom. Senzitivnost je važan statistički parametar za potencijalnu primenu prediktivnih alata jer meri stopu lažno negativnih predikcija. Fišerov test je pokazao da je ISM algoritam statistički značajan za predviđanje funkcionalnih efekata nCFD pozicioniranih AAS u epigenetičkim regulatorima ($p < 0,001$).

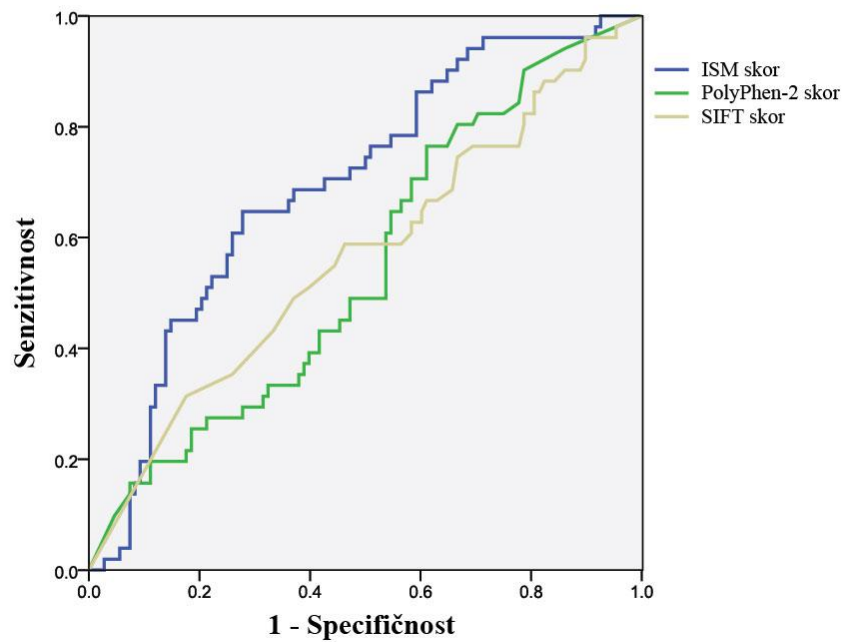
Tabela 15. Statistički parametri efikasnosti predviđanja funkcionalnih efekata aminokiselinskih supstitucija ISM algoritma na nCFD_AAS skupu (p vrednosti su dobijene Fišerovim testom).

	TP	TN	LP	LN	Senzitivnost	Specifičnost	Tačnost
ISM algoritam ($p < 0,001$)	33	76	32	18	0,65	0,70	0,69

LN – lažno negativne predikcije; LP – lažno pozitivne predikcije; TN – tačno negativne predikcije; TP – tačno pozitivne predikcije.

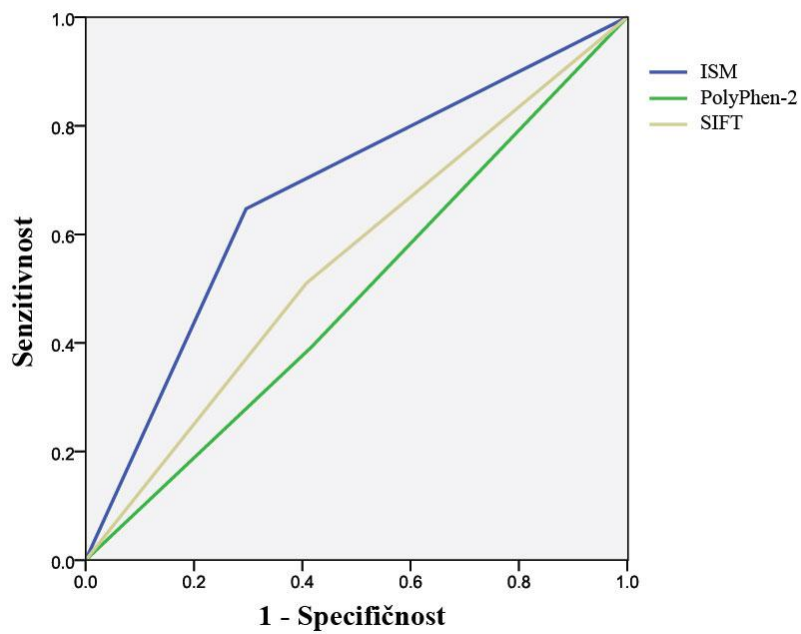
Za poređenje ISM algoritma sa PolyPhen-2 i SIFT je, takođe, korišćena ROC analiza, pri čemu su poređene: i) vrednosti skorova dobijenih ovim alatima i ii) njihove MUT/SNP predikcije. Analiza skorova dobijenih pomoću PolyPhen-2, SIFT i ISM algoritama, metodom ROC krive, je pokazala najbolju efikasnost ISM alata, sa vrednostima AUC 0,55 za PolyPhen-2, 0,57 za SIFT i 0,70 za ISM algoritam ([Slika 21](#)). Ovaj rezultat je potvrđen i analizom samih MUT/SNP predikcija. ISM predikcije su pokazale veću vrednost AUC za 19% u odnosu na PolyPhen-2 i 13% u odnosu na SIFT ([Slika 22](#)).

ROC kriva



Slika 21. ROC analiza PolyPhen-2, SIFT i ISM skorova kao prediktora funkcionalnih efekata aminokiselinskih supstitucija pozicioniranih u nCFD.

ROC kriva



Slika 22. ROC analiza PolyPhen-2, SIFT i ISM predikcija funkcionalnih efekata aminokiselinskih supstitucija pozicioniranih u nCFD.

5. Diskusija

Od svih tipova maligniteta, leukemije su na jedanaestom mestu po incidenci, sa oko 352000 novodijagnostikovanih pacijenata u 2012. godini (Ferlay i sar.2013), a smatra se da će se njihova incidenca povećavati u skladu sa starenjem ljudske populacije. Sadašnja prosečna petogodišnja stopa preživljavanja je manja od 40%, zbog čega je neophodan razvoj novih terapijskih pristupa. Svi tipovi mijeloidnih maligniteta se karakterišu velikom intra- i interkancerskom genetičkom heterogenošću, pa postoji potreba za individualizovanim pristupom svakom pacijentu. Bioinformatika ima važnu ulogu u personalizovanoj medicini, u identifikaciji ciljnih molekula za razvoj terapije, racionalnom dizajnu lekova i detekciji informativnih biomarkera za prateću dijagnostiku.

Ekperimentalne tehnike za identifikaciju PPI su izuzetno skupe, a, pored toga, pokazuju neke nedostatke, npr. nisku pokrivenost interakcija i tendenciju ka određenim tipovima proteina i ćelijskoj lokalizaciji (Shoemaker i Panchenko 2007). Zbog toga je neophodno razvijati nove tehnike za kompjuterske predikcije PPI, koje će na osnovu postojećeg znanja i podataka omogućiti formulisanje hipoteza, koje bi se dalje eksperimentalno testirale. To bi smanjilo troškove i obuhvatilo sve tipove proteina i interakcija. Savremene tendencije u predviđanju PPI podrazumevaju integrisanje informacija iz različitih izvora, kao što su sekvence proteina, njihove anotacije i literaturni podaci (v. Poglavlje 1.3.2.). U okviru ove disertacije je korišćen skup koraka za predikciju PPI, zasnovan na informacijama o primarnoj sekvenci proteina, funkcionalnim anotacijama, mreži PPI i ekspresiji proteina u određenom tkivu/ćelijskoj liniji. Specifičnost ovog algoritma je upotreba ISM metode za analizu primarne sekvence proteina. Ova metoda, kao i BLAST ili PSI-BLAST (Altschul i sar. 1997), omogućava pretraživanje velikih baza sekvenci sa ciljem identifikacije sekvenci sličnih ulaznoj. Međutim, sličnost sekvenci se ne detektuje na osnovu pozicije određenih aminokiselina na određenim mestima u sekvenci, kao što je slučaj sa MSA-zasnovanim analizama, već na osnovu neočiglednih strukturnih obrazaca identifikovanih Furijeovom transformacijom. Obrasci detektovani ISM-om su relevantni za predikciju PPI jer ova metoda koristi EIIP, elektrodinamičku komponentu dalekosežnih međumolekulskih interakcija, kao parametar za kodiranje proteinskih sekvenci.

WT1 je multifunkcionalni protein. Njegova najvažnija uloga je u transkripcionoj regulaciji ekspresije gena uključenih u apoptozu, proliferaciju i diferencijaciju. WT1 ima ulogu i u post-transkripcionom metabolizmu iRNK, u splajsovanju (Larsson i sar. 1995) i vezivanju određenih formi iRNK za polizome (Bor i sar. 2006). Funkcije WT1 su zavisne od konteksta, tj. modulirane su velikim brojem kofaktora (Roberts 2006). Za identifikaciju novih PPI proteina WT1 primenjen je integrativni algoritam. Analiza informacionih spektara različitih izoformi ovog proteina, kao i sekvenci iz različitih sisarskih vrsta, pokazala je da je dominantna frekvenca u svim ovim slučajevima ista, i to je – F(0.00195). Daljim skeniranjem UniProt baze kros-spektralnom analizom, detektovani su proteini kod kojih je intenzitet (reprezentovan S/N vrednošću) ove frekvence u spektru sa WT1 iznad granične vrednosti. Na taj način su izdvojeni proteini koji učestvuju u istim/sličnim biološkim procesima kao i WT1, potencijalno formirajući komplekse sa ovim proteinom. *Enrichment* analizom ovog skupa proteina se pokazalo da su oni, sa najvećom značajnošću, uključeni u procese transkripcije i obrade iRNK, što je u saglasnosti sa utvrđenim najvažnijim ulogama WT1 (npr. Hohenstein i Hastie 2006). Ovaj rezultat pokazuje opravdanost primene ISM u predviđanju funkcionalno značajnih interakcija. Takođe, to što se pokazalo da je ista frekvenca asocirana i sa transkripcijom i sa obradom iRNK, u saglasnosti je sa otkrićima drugih autora (Ladomery 1997; Bentley 2002; Nogués i sar. 2003; Kornbliht i sar. 2004).

Integriranjem ISM, GO i STRING rezultata, dobijena je lista 51 kandidata za nove interaktore WT1, a sledeći korak je njihova eksperimentalna validacija. S obzirom da je WT1 endogeno eksprimiran u K562 ćelijskoj liniji, ona predstavlja često korišćen model za ispitivanje PPI proteina WT1 (Lee TH i sar. 2002; Carpenter i sar. 2004; Morrison AA i Laodmery 2006; Makki i sar. 2008; Green i sar. 2009; Bansal i sar. 2010; Florio i sar. 2010; Hartkamp i sar. 2010). U poslednjem koraku integrativnog algoritma za predikciju PPI proteina WT1, dobijeno je 36 proteina koji su: i) sa zajedničkim spektralnim svojstvima sa WT1; ii) uključeni u regulaciju transkripcije; iii) u najbližem susedstvu WT1 u mreži humanih PPI; iv) koeksprimirani sa WT1 u K562 ćelijama.

Najveći broj proteina koji su identifikovani u ovom radu kao kandidati za nove interaktore WT1 su transkripcioni faktori, koji regulišu ekspresiju gena uključenih u ključne ćelijske procese – proliferaciju, diferencijaciju i apoptozu. To su: CTNNB1 (β -katenin) (Lustig i Behrens 2003), EGR1 (Thiel i Cibelli 2002), MEF2C (Edmondson i sar. 1994; Li Z i sar. 2008; Stehling-Sun i sar. 2009), SNIP1 (Kim RH i sar. 2000, 2001; Fujii i sar. 2006), SRF (Miano 2003), VEGFA (Ferrara i sar. 2003) i YY1 (Gordon i sar. 2006). Kandidati su i dva proteina, sa funkcijom interaktora i modulatora transkripcionih faktora, pre svega, nuklearnih receptora: DDX54 (Rajendran i sar. 2003) i LCOR (Asim i sar. 2011).

Potencijalni interaktori WT1 su i proteini uključeni u odgovor ćelije na stres i virusnu infekciju: IRF1 (Fujita i sar. 1989), KEAP1 (Zhang DD i Hannink 2003; Kobayashi i sar. 2004), KHDRBS1 (SAM68) (Bielli i sar. 2011), NFATC2 (Peng SL i sar. 2001), POU2F1 (OCT-1) (Tantin i sar. 2005) i STAT2 (Leung S i sar. 1995; Bhattacharya i sar. 1996).

Skup novih PPI proteina WT1 obuhvata i tri ciklin-zavisne kinaze (engl. *Cycline Dependant Kinases*, CDK) – CDK2, CDK8 i CDK9. CDK su katalitičke jedinice velike porodice heterodimernih serin/treonin proteinskih kinaza, koje za svoju aktivnost zahtevaju vezivanje ciklinske regulatorne komponente (Malumbres i Barbacid 2005). Ovi kompleksi su uključeni u kontrolu progresije ćelije kroz ćelijski ciklus (CDK2) i transkripcije (CDK8 i CDK9).

Ekspresija gena je regulisana epigenetičkim modifikacijama, koje obuhvataju modifikacije DNK (metilaciju/demetilaciju) i različite hemijske modifikacije histona. U grupi od 36 predviđenih novih interaktora WT1, sedam proteina je uključeno u epigenetičke procese. DNMT3B i DMAP1 učestvuju u procesu metilacije DNK, dok RING1, KAT2B, KDM1A, MTA1 i AEBP2 učestvuju u hemijskom modifikovanju histona. DNMT3B učestvuje, zajedno sa DNMT3A, u *de novo* metilaciji DNK (Okano i sar. 1999). A pokazano je da ovaj enzim, u kooperaciji sa DNMT1, DNK metiltransferazom odgovornom za održavanje metilacije DNK, održava metilaciju DNK u humanim malignim ćelijama. DNMT1/DNMT3B metilacija DNK vodi utišavanju genske

ekspresije nekih tumor supresorskih gena, pa je tako uključena u neoplastičnu proliferaciju ćelija (Rhee i sar. 2002). DMAP1 interaguje sa DNMT1 i učestvuje u epigenetičkoj represiji genske ekspresije (Rountree i sar. 2000). WT1, takođe, direktno interaguje sa DNMT1 (Xu B i sar. 2011). Xu i sar. su pokazali da WT1 regrutuje DNMT1 na promotor PAX2 gena, što dovodi do hipermetilacije promotora i represije ovog gena. RING1 je ubikvitin-protein ligaza, koja vrši ubikvitinaciju histona H2A (Stock i sar. 2007). Ovaj protein je član PRC1 represorskog kompleksa (Satijn i sar. 1997). KAT2B je histon acetiltransferaza (Jin Q i sar. 2011). KDM1A i MTA1 su komponente NuRD (engl. *Nucleosome Remodeling Deacetylase*) kompleksa odgovornog za deacetilaciju histona i remodelovanje hromatina, što vodi transkripcionoj represiji (Xue i sar. 1998; Wang Y i sar. 2009). AEBP2 je transkripcioni faktor, koji interaguje sa PRC2 kompleksom, ključnim za metilaciju histona, i na taj način učestvuje u epigenetičkoj regulaciji genske ekspresije (Cao i Zhang 2004; Kim H i sar. 2009). WT1, takođe, interaguje sa komponentama PRC2 kompleksa – EZH2 i SUZ12, u regulaciji ekspresije PAX2 gena (Xu B i sar. 2011). Identifikacija ovih sedam epigenetičkih regulatora kao potencijalno novih interaktora WT1 ukazuje na ulogu WT1 u regulaciji epigenetičkih procesa i na značaj WT1 regulacije represorskih PRC1 i PRC2 kompleksa.

Sinteza iRNK u eukariotskim ćelijama je katalizovana enzimskim kompleksom RNK polimeraze II (PolII). Ovaj proces je kritično mesto kontrole genske ekspresije, regulacija se odigrava i u fazi inicijacije i u fazi elongacije transkripcije (pregled u Conaway i Conaway, 2011). Kontrolu regrutovanja PolII na promotore ciljnih gena i inicijacije i elongacije transkripcije vrši veliki broj transkripcionih regulatora. Jezgro ovog skupa proteina čine evoluciono konzervirani transkripcioni faktori neophodni za funkcionisanje PolII: i) opšti faktori inicijacije transkripcije GTF2B, GTF2D, GTF2E, GTF2F, GTF2H; ii) faktori elongacije transkripcije GTF2S, ELL, DSIF, Elongin i P-TEFb; iii) Medijatorski kompleks, koga čini veliki broj proteina, uključujući MED12, MED13, MED26 i CDK8. Sedam predviđenih novih interaktora WT1 učestvuje u regulaciji aktivnosti PolII. U inicijaciji transkripcije učestvuju GTF2E1 (subjedinica 1 opšteg transkripcionog faktora TFIIE), subjedinice faktora inicijacije GTF2D – TAF3, TAF5 i TAF10, CDK9 (katalitička

subjedinica P-TEFb) i komponente Medijatorski kompleksa – MED15 i CDK8. Do sada je opisana interakcija WT1 samo sa dva opšta transkripciona faktora uključena u inicijaciju transkripcije – GTF2D1 i GTF2B (McKay i sar. 1999; Richard i sar. 2001), tako da ovde definisana mreža u velikoj meri doprinosi objašnjenju mehanizma delovanja WT1 na PolII.

Iako nije deo PolII mašinerije, predviđeni interaktor SAFB utiče na regrutovanje ovog kompleksa na promotore gena (Nayler i sar. 1998). SAFB vezuje S/MARs (engl. *Scaffold/Matrix Attached Regions*) hromatina, koji su locirani u blizini granica gena u aktivnoj transkripciji i utiču na njihovu aktivnost. Ovaj protein, takođe, vezuje PolII mašineriju i neke faktore splajsovanja, pa se pretpostavlja da predstavlja bazu za grupisanje ovih proteina na S/MAR regione hromatina na granicama gena u aktivnoj transkripciji. CTCF učestvuje u strukturnoj organizaciji hromozoma, imprintingu i inaktivaciji X hromozoma (Phillips i Corces 2009). A pokazano je da CTCF reguliše splajsovanje iRNK delujući na brzinu transkripcije kroz interakciju sa PolII, pri čemu je ova funkcija CTCF modulirana epigenetičkim modifikacijama DNK (Shukla i sar. 2011).

Četiri kandidata za PPI proteina WT1 su lokalizovani u splajsozomu i uključeni u splajsovanje - HNRNPUL1 (E1B-AP5), PUF60 (FIR), RBM39 i SF1. Za PUF60, RBM39 i SF1 je pokazano da interaguju sa U2AF65 faktorom splajsovanja (Selenko i sar. 2003; Prigge i sar. 2009), poznatim interaktorom WT1 (Davies i sar. 1998). Interakcija između SF1 i U2AF65 je neophodna za prepoznavanje mesta splajsovanja na pre-iRNK, što je prvi korak u generisanju splajsozoma (Selenko i sar. 2003).

Svi ovi koraci u regulaciji ekspresije gena – epigenetičke modifikacije, transkripcija i splajsovanje, su povezani, prostorno i vremenski, u velike komplekse 'iRNK fabrike' (McCracken i sar. 1997; Bentley 2002; Nogués i sar. 2003; Kornbliht i sar. 2004; Zorio i Bentley 2004; Bentley 2005; Hagiwara i Nojima 2007; Stock i sar. 2007; Muñoz i sar. 2010; Shukla i sar. 2011). Kombinacija ISM i *enrichment* analize je pokazala da proteini koji interaguju sa WT1, a uključenu su u ova tri procesa imaju zajednička spektralna svojstva – $F(0.00195)$. S obzirom da su spektri ovih proteina zasnovani na EIIP, dinamičkoj komponenti dalekosežnih međumolekulskih interakcija, ovi rezultati mogu poslužiti kao

objašnjenje mehanizma međusobnog prepoznavanja proteina prilikom formiranja (engl. *assembly*) 'iRNK fabrika'. Iako su već postojali nagoveštaji da WT1 igra važnu ulogu u regulaciji 'iRNK fabrika', ovde identifikovana mreža potencijalnih PPI ovog proteina potvrđuje te pretpostavke i u značajnoj meri upotpunjuje sliku o mehanizmima ove regulacije.

Interakcije specifičnih parova ili grupa proteina su ključne za sve faze razvoja i homeostaze. Narušavanje ovih PPI, bilo kroz gubitak interakcije ili formiranje neodgovarajućih kompleksa, leži u osnovi mnogih bolesti, uključujući i karcinome. Normalne ćelije prelaze u maligno stanje sticanjem sposobnosti samoodržive proliferacije, izbegavanja supresora rasta, suprotstavljanja ćelijskoj smrti i sticanjem replikativne besmrtnosti, a neki karcinomi, pored toga, indukuju angiogenezu i dovode do nastanka metastaza (Hanahan i Weinberg 2000). Aberantne PPI omogućavaju ćeliji sticanje svakog od ovih svojstava, pa terapijsko delovanje na njih ima veliki antikancerski potencijal. Na primer, interakcija između AKT i mTOR proteina, u slučaju aktivacije PI3K/AKT/mTOR signalnog puta mutacijom ili amplifikacijom, dovodi do povećane sinteze proteina koji omogućavaju samoodrživu proliferaciju (npr. c-Myc) (Hennessy i sar. 2005; LoPiccolo i sar. 2008). Fizičke interakcije između pro-apoptotskih članova porodice BCL-2 proteina (BAX, BAK, BIM) i anti-apoptotskih BCL-2 i BCL-X_L regulišu ulazak i održavanje ćelije u apoptozi. Zato je važan mehanizam suprotstavljanja ćelijskoj smrti malignih ćelija povećana ekspresija anti-apoptotskih ili smanjena ekspresija pro-apoptotskih gena iz ove grupe (Adams i Cory 1998; Mukherjee i sar. 2010). Neke PPI omogućavaju karcinomima sticanje većeg broja svojstava, pa su takve interakcije glavni kandidati za mete za delovanje ciljane, na PPI usmerene, terapije. TP53 protein ima ulogu u regulaciji ćelijskog ciklusa, popravci oštećenja DNA, angiogenezi i apoptozi, a njegova aktivnost je direktno regulisana interakcijom sa proteinom MDM2 (Momand i sar. 1992). Inhibitori ove PPI su u različitim fazama prekliničkih i kliničkih ispitivanja (Vassilev i sar. 2004; Lu i sar. 2006; Lawrence i sar. 2009; Andreef i sar. 2010; Rew i sar. 2012; Zhuang i sar. 2012; Brown CJ i sar. 2013).

WT1, u određenim kontekstima, u malignim ćelijama izaziva njihova dva ključna svojstva – povećanu proliferaciju (Algar i sar. 1996; Yamagami i sar. 1996; Tsuboi i sar. 1999; Han i sar. 2004; Tuna i sar. 2005; Gao SM i sar. 2011; Xu C i sar. 2013; Kijima i sar. 2014; Rather i sar. 2014) i izbegavanje apoptoze (Maheswaran i sar. 1995; Algar i sar. 1996; Mayo i sar. 1999; Tuna i sar. 2005; Ito K i sar. 2006; Tatsumi i sar. 2008; Uesugi i sar. 2013; Watanabe i sar. 2013; Kijima i sar. 2014). Ovo je posledica uloge WT1 u regulaciji transkripcije određenih gena, kao što su c-Myc, vitamin D, gena Bcl-2 porodice i drugih. Ekspertskom analizom liste potencijalna PPI proteina WT1, utvrđeno je da je najrelevantniji kandidat za ciljanu terapiju, interakcija WT1 i CDK9. Ova PPI je, nakon predikcije u ovom istraživanju, eksperimentalno i potvrđena na Odseku za biološke nauke Univerziteta u Bufalu, SAD (Roberts i sar. rad u pripremi). Koimunoprecipitacija, na nuklearnim ekstraktima K562 ćelija, je pokazala da anti-WT1 antitela precipitiraju CDK9 i, obrnuto, anti-CDK9 antitela su precipitirala WT1. Ovaj rezultat se može smatrati značajnom potvrdom efikasnosti integrativnog algoritma za predikciju PPI prikazanog u ovom radu.

CDK9 je subjedinica P-TEFb (engl. *Positive Transcription Elongation Factor b*) kompleksa, koji je ključni element PolII mašinerije. Ovaj kompleks je neophodan za elongaciju transkripcije, tj. za oslobađanje PolII iz pauze u produktivnu elongaciju (Price 2000; Peterlin i Price 2006). P-TEFb kompleks se sastoji iz katalitičke kinazne subjedinice CDK9 i jednog od tri regulatorna ciklina – T1, T2 ili K (Peng J i sar. 1998; Fu i sar. 1999). RNApolIII je pozicionirana na promotorima nekih gena i pre početka transkripcije, ali u pauzi. P-TEFb aktivira elongaciju izvlačeći PolII iz pauze tako što stupa u interakciju i fosforiliše njen C-terminalni domen (CTD). Ovakav model transkripcije omogućava brzi transkripcioni odgovor na spoljašnji stimulus i karakterističan je za gene uključene u razvoj i odgovor na stres. Na kojim genima će doći do aktivacije transkripcije zavisi od regrutovanja P-TEFb na promotor, tj. od repertoara transkripcionih faktora sposobnih da vežu P-TEFb za promotor gena. CIITA transkripcioni faktor ima sposobnost regrutovanja P-TEFb na promotore gena za MHC klase II (Kanazawa i sar. 2000), NF-kB regrutuje P-TEFb na promotor gena za IL-8 (Barboric i sar. 2001; Nowak i sar. 2008), a c-Myc na

promotor gena za CAD (Eberhardy i Farnham 2001, 2002; Kanazawa i sar. 2003; Rahl i sar. 2010).

Za WT1 je ranije pokazano da reguliše prisustvo RNAPolIII na promotoru Wnt4 gena. Esaffi i saradnici su pokazali da tkivno specifične PPI proteina WT1 utiču na pauziranje RNAPolIII mašinerije i aktivaciju transkripcione elongacije (Esaffi i sar. 2011). Interakcija između WT1 i CDK9 bi mogla da objasni mehanizam ovakvog delovanja WT1. U bubrezima, WT1 vezuje CBP/p300 kofaktor, što dovodi do izlaska RNAPolIII iz pauze i aktivacije elongacione faze transkripcije. Gubitak WT1 u bubrezima vodi gubitku fosforilacije CTD PolIII. S obzirom da CDK9 fosforiliše CTD PolIII, ovo bi se moglo objasniti gubitkom interakcije između CDK9 i aktivirajućeg WT1/CBP/p300 kompleksa na promotoru Wnt4 gena. S druge strane, u epikardiumu, WT1 vezuje BASP1 kofaktor, što dovodi do pauziranja RNAPolIII mašinerije i represije Wnt4 gena. Gubitak WT1 u epikardiumu vodi povećanju fosforilacije CTD PolIII, što bi se moglo objasniti gubitkom interakcije CDK9 i WT1/BASP1 represorskog kompleksa na promotoru Wnt4 gena. Dakle, ovde dobijeni rezultati upućuju na to da WT1 (u kompleksu sa tkivno specifičnim kofaktorima) reguliše fosforilaciju CTD PolIII i, posledično, elongaciju transkripcije, interakcijom sa CDK9.

CDK9 je jedan od ključnih targeta u antikancerskoj ciljanoj terapiji, a ima važnu primenu i u hematologiji (pregled u Wang S i Fischer 2008; Krystof i sar. 2012). CDK inhibitori, kao što su flavopiridol, roskovitin i AT7519, deluju na ćelijski ciklus kroz inhibiciju CDK1 i 2, ali je glavni target za njihovo antionkogeno delovanje CDK9 i regulacija transkripcije gena uključenih u kancerogenezu (Cai i sar. 2006). CDK9, u interakciji sa transkripcionim faktorima, npr. c-Myc, reguliše transkripciju velikog broja gena uključenih u proliferaciju i apoptozu (Kanazawa i sar. 2003; Rahl i sar. 2010). Za neke ključne gene iz ovog skupa, koji imaju ulogu u patogenezi maligniteta, je pokazano da je ekspresija regulisana proteinom WT1. To je slučaj sa pro-apoptotskim Mcl-1 (Wang X i sar. 2013), koji je glavni efektor CDK9-posredovane apoptoze (Gojo i sar. 2002; MacCallum i sar. 2005; Raje i sar. 2005; Santo i sar. 2010; Polier i sar. 2011). Ovi podaci ukazuju na mogućnost da je bar deo terapeutskih efekata CDK inhibitora posledica blokade

transkripcije gena koja je posredovana WT1-CDK9 interakcijom, što znači da je ona potencijalno dobar target za ciljanu terapiju.

Za dizajn malih molekula sa terapeutskim delovanjem na ovu PPI, neophodno je poznavanje interagujućih domena ova dva proteina, pa je značajno što je CDK9 struktura kristalografski određena, uključujući i strukturu u interakciji sa inhibitorima (Brown NR i sar. 1999; Baumli i sar. 2008; 2010; Tahirov i sar. 2010). Takođe, u proteklih godinu dana su *in silico* dizajnirana jedinjenja koja inhibiraju interakciju CDK9 i ciklina T1 (Randjelovic i sar. 2014). Ova znanja se mogu iskoristiti za inhibiciju i drugih interakcija CDK9, uključujući potencijalno i onu sa WT1. Treba istaći činjenicu da je WT1 specifično eksprimiran u malignim ćelijama (Miwa i sar. 1992; Menssen i sar. 1995; Inoue i sar. 1997; Ostergaard i sar. 2004), pa bi efekti inhibitora bili specifični i sa malo neželjenih dejstava na normalne ćelije.

NISCH, trenutno jedini kandidat za I₁ imidazolinski receptor (v. Poglavlje 1.2.2.1.), je novi tumor supresor u karcinomu dojke (Ding i sar. 2008, Baranwal i sar. 2011). Tumor supresorsku ulogu NISCH ispoljava preko RAC signalnog puta, koji je ključan u regulaciji hematopoeze i uključen je u patogenezu mijeloidnih maligniteta (Skorski i sar. 1998; McGee i sar. 2002; Harnois i sar. 2003; Gu i sar. 2003; Somerville i Cleary 2006; Thomas EK i sar. 2007, 2008; Müller i sar. 2008; Cancelas i Williams 2009; Karlsson i sar. 2009; Hinterleitner i sar. 2013; Martin i sar. 2013). Zbog toga je ovaj protein novi potencijalni target za terapeutsko delovanje u mijeloidnim malignitetima.

Uloga NISCH-a u apoptozi je kompleksna i zavisna od konteksta (Dontenwill i sar. 2003; Dupuy i sar. 2004; Aceros i sar. 2011). Nakon što smo pokazali da rilmenidin, agonist I1-IR, može da izazove apoptozu u leukemijskoj ćelijskoj liniji (Nikolic i sar. 2013), cilj nam je bio da predvidimo potencijalno mesto vezivanja. Ovaj podatak bi se u budućnosti koristio za izbor novih kandidata za lekove.

U prvom koraku ovog dela istraživanja su identifikovane dve dominantne frekvence u informacionom spektru proteina NISCH – F(0,152) i F(0,426). Iterativnim isecanjem prozora od 55 do 60 ak iz sekvence NISCH-a, generisan je skup novih sekvenci. Na osnovu

spektralne analize utvrđeno je da intenzitet lokalnog maksimuma na F(0,152) maksimalno smanjuje isecanje polipeptida koji obuhvata region od 680 do 739 rezidue. Na isti način je identifikovano da isecanje regiona od pozicije 966 do 1020 maksimalno smanjuje intenzitet lokalnog maksimuma na F(0,426). Delecija ovog domena dovodi do potpunog pada amplitude na frekvenci F(0,426), približno nuli, što je dodatno ukazalo na važnost ovog domena za funkciju NISCH-a.

Strukturna analiza je pokazala da se region 680-739ak delimično preklapa sa *coiled-coil* regionom odgovornim za homooligomerizaciju ovog proteina (Lim i Hong 2004). Identifikovani biološki procesi u kojima učestvuje ovaj domen NISCH-a se mogu grupisati u dve grupe: i) uključeni u sinaptogenezu i ii) uključeni u PTM, specifično ubikvitinaciju. Uloga NISCH-a u funkcionisanju nervnih ćelija je opisana i ranije. NISCH učestvuje u prenosu signala sa EphB2, tirozin kinaznog receptora uključenog u razvoj nervnog sistema (Zhang G i sar. 2006; Darie i sar. 2011). Ovaj signalni put je važan za formiranje nervne kreste i sinapsi (Holland i sar. 1996; Robinson i sar. 1997; Holland i sar. 1998; Dalva i sar. 2000; Wilkinson 2001). NISCH je lociran u perinuklearnim regionima neurona i protruzijama bogatim F-aktinom u mozgu pacova, gde reguliše migraciju neurona (Ding i sar. 2013). Signalni put kojim NISCH reguliše ćelijske migracije u karcinomima dojke i kolona uključuje njegove PPI sa $\alpha 5$ integrinskom subjedinicom fibronektinskog receptora (ITGA5) i PAK i LIMK1 kinazama RAC signalnog puta (Alahari i sar. 2000, 2004; Alahari 2003; Ding i sar. 2008; Baranwal i sar. 2011). Može se pretpostaviti da NISCH kroz ovaj signalni put reguliše i funkcije i migraciju neurona jer je PAK4/LIMK1 signalni put ključni regulator rasta aksona/dendrita i migracije neurona (Eswaram i sar. 2008). Dakle, navedeni literaturni podaci pokazuju da je *in silico* funkcionalna karakterizacija domena 680-739ak proteina NISCH, prikazana u ovoj disertaciji, u saglasnosti sa utvrđenom ulogom u regulaciji funkcija neurona i formiranja sinapsi. Posebno značajna potvrda ovde korišćene metodologije za funkcionalnu karakterizaciju NISCH-a je to što je ovaj region (i njegovu okolinu u sekvenci NISCH-a) u UniProt bazi anotiran, na osnovu manuelnog pregledanja sličnosti sekvenci, kao važan za interakciju sa ITGA5, PAK1 i LIMK1 proteinima (UniProt Consortium 2014).

Pored funkcije u sinaptogenezi i razvoju nervnog sistema, koja je prethodno pokazana za NISCH, ovo istraživanje je ukazalo i na PTM i ubikvitinaciju kao važne biološke procese asocirane sa 680-739ak domenom. Ovaj rezultat može da se interpretira na dva načina: i) PTM/ubikvitinacija je uključena u funkcionisanje i/ili proteolizu NISCH-a ili ii) NISCH ima ulogu u PTM/ubikvitinaciji drugih proteina. Ova druga funkcija je manje verovatna jer u sekvenci NISCH-a nisu identifikovani regioni koji bi imali katalitičku ulogu (Piletz i sar. 2000). Darie i sar. ukazuju na mogućnost PTM proteina NISCH, pre svega fosforilacije (Darie i sar. 2011). Ova fosforilacija NISCH-a je uključena u EphB2 signalni put, koji je, kao što je već napomenuto, važan za sinaptogenezu i razvoj nervnog sistema. Dakle, rezultati u ovoj disertaciji pokazuju da su za PTM NISCH-a i njegovu ulogu u funkcionisanju nervnih ćelija odgovorna ista spektralna svojstva, što može objasniti mehanizam delovanja NISCH-a u EphB2 signalnom putu. Funkcionalno asociranje regiona 680-739ak sa ubikvitinacijom, obeležavanjem proteina ubikvitinom za degradaciju u proteazomu (Ciechanover 1994), implicira da bi NISCH mogao biti substrat proteazomalnog puta degradacije, što, do sada, nije eksperimentalno testirano, mada postoje preliminarni rezultati koji bi bili u skladu sa tim (Bair i sar. 2008).

Do sada nema istraživanja čiji rezultati ukazuju na biohemijsku ulogu domena 966-1020ak NISCH-a. Funkcionalna anotacija izvedena u ovom radu predviđa tri klastera bioloških procesa za ovaj domen NISCH-a: i) apoptoza, ii) transkripcija i iii) signalna transdukcija posredovana hormonima. Apoptoza je anotirana u GO kao proces u koji je uključen NISCH, a ovde je prvi put detektovan domen koji posreduje u ovoj funkciji NISCH-a. Takođe, anotacija NISCH-a u GO podrazumeva i neke puteve signalne transdukcije, posredovane GPCR i RAC proteinima, ali ne i regulaciju transkripcije.

Endogeni ligandi IR su agmatin, IAA-RP (engl. *imidazol-4-acetic acid ribotide*) i harmani i harmalani (Nikolić i Agbaba 2012). Sintetički ligandi I₁-IR se mogu podeliti u tri kategorije: i) agonisti – rilmenidin, benazolin i moksonidin, ii) delimični agonisti – efaroksan i iii) antagonisti – S23757, LNP906 i LNP911. S obzirom na to da ligandi imidazolinskih receptora pokazuju dobru selektivnost ka IR, njihove strukture predstavljaju veoma dobru osnovu za razvoj novih aktivnijih i selektivnijih imidazolinskih liganada. Na

primer, otkriće da NISCH interaguje sa IRS-4 u insulinskom signalnom putu je dovelo do ispitivanja novog I₁-IR agoniste S43126 kao potencijalnog leka za insulinsku rezistenciju u dijabetesu tipa 2 (Tesfai i sar. 2011). Za NISCH je, takođe, pokazano da posreduje u efektima agmatina, endogenog liganda I₁-IR u mozgu, na razvoj morfinske zavisnosti, pa je predložen za novi target za razvoj terapije za lečenje morfinske zavisnosti (Li F i sar. 2012). S obzirom da su I₁-IR identifikovani na ćelijama srca i da imaju efekat na preživljavanje/apoptozu ovih ćelija, što zavisi od tipa srčanih ćelija, predloženo je da bi novi lekovi sa svojstvima specifičnim za tip ćelija mogli biti korišćeni za sprečavanje razvoja srčane insuficijencije, kao dobar dodatak postojećoj terapiji (Mukaddam-Daher 2012). Nakon identifikacija 966-1020ak regiona NISCH-a, metodom informacionih spektara, kao potencijalnog mesta vezivanja liganada, usledile su i *virtual docking* i 3D-QSAR studije. Ovo je rezultovalo utvrđivanjem modela vezivanja I₁-IR agonista i antagonista (Nikolic i sar. 2013). Primena ovih rezultata je u racionalnom dizajnu i evaluaciji novih jedinjenja, koja bi mogla imati terapeutske efekte u malignitetima u čijem nastanku učestvuje NISCH – karcinom dojke, kolona i CML.

Drugi klaster pojmova za funkcionalnu anotaciju NISCH domena 966-1020ak, određen ISM analizom, je regulacija transkripcije i obuhvata transkripcione faktore sa domenima cink-prstića. Iako ovaj biološki proces još uvek nije asociiran sa NISCH-em u GO bazi, postoje literaturni podaci koji idu u prilog ulozi ovog proteina u transkripciji:

- Baranwal i sar. su pokazali da NISCH transkripciono reguliše ekspresiju ITGA5 gena, *in vitro* i *in vivo* (Baranwal i sar. 2011). NISCH na ovaj način ispoljava tumor supresorski efekat u ćelijama karcinoma dojke.
- NISCH inhibira transkripcionu aktivaciju promotora gena ciklin D1, posredno kroz interakciju sa mutiranim RAC1 proteinom (Reddig i sar. 2005). Mutirani RAC1 dovodi do aktivacije NF-κB i ekspresije ciklin D1 gena, što je inhibirano delovanjem NISCH-a. Supresija ekspresija gena za ciklin D1, kao i gena CDK4, mogu biti posredovani i interakcijom NISCH-a sa LKB1, efektorom RAC1 signalnog puta (Jain i sar. 2013).
- NISCH reguliše sopstvenu ekspresiju putem efekata na NF-κB signalni put (Jin L i sar., 2013). Aktivacija NF-κB dovodi do povećane ekspresije mikro RNK miR-

23b/27b, što vodi smanjenoj ekspresiji NISCH. U ćelijama sa povećanom ekspresijom NISCH-a dolazi do supresije NF-kB puta, koja je praćena smanjenim prisustvom fosforilisanih p65 i ERK, a povećanim prisustvom I κ B α , što za posledicu ima smanjenu ekspresiju miR-23b/27b i delovanje povratnom spregom na ekspresiju samog NISCH-a.

- Uticaj NISCH/IRAS na transkripciju je zabeležen i u CHO ćelijama koje nemaju endogeni I₁-IR, već su transfekovane humanim NISCH/IRAS (Wu N i sar. 2006). CHO ćelije dugotrajno tretirane morfinom predstavljaju model za morfinску zavisnost. Naknadno tretiranje ovih ćelija naloksonom, I₁-IR ligandom, dovodi do povećane fosforilacije CREB i ERK1/2 i posledično povećane genske ekspresije c-Fos. Svi ovi efekti su blokirani zajedničkim pretretmanom ćelija morfinom i agmatinom, endogenim ligandom I₁-IR.
- NISCH utiče i na ekspresiju gena BCL-2 porodice, uključenih u apoptotski put neurona, kardiomiocita i fibroblasta (Wu X i sar. 2013; Aceros i sar. 2014).

Rezultati ovog rada su u saglasnosti sa literaturnim podacima i pokazuju da ISM algoritam može imati značajnu primenu u predviđanju bioloških procesa čak i za slabo funkcionalno anotirane proteine i njihove domene. Iako postoje eksperimentalni pokazatelji da NISCH učestvuje u procesu regulacije transkripcije, mehanizmi kojima se postiže ovaj efekat nisu poznati. U ovom radu je prvi put izveden rezultat koji povezuje NISCH i jedan deo njegove sekvence sa ovim biološkim procesom. Poznavanje frekventnog domena u informacionom spektru NISCH-a, koji je asociran sa ovom funkcijom (F(0,426)), može da doprinese daljem rasvetljavanju molekulskih mehanizama, npr. detekcijom interakcija proteina NISCH.

Najveći broj bioinformatičkih metoda za predviđanje štetnih efekata polimorfizama se zasniva na analizi sekvence i strukture proteina, a glavna pretpostavka je da se AAS asocirane sa bolestima nalaze na evoluciono konzerviranim pozicijama u proteinskoj sekvenci (Jordan i sar. 2010). Rezultati ovog istraživanja pokazuju da najčešće korišćeni alati, PolyPhen-2 i SIFT, zaista, sa velikom tačnošću predviđaju mutacije u CFD epigenetičkih regulatora ASXL1, EZH2, DNMT3A i TET2. PolyPhen-2 je tačno predvideo

za 97.2%, a SIFT za 90.2% AAS u CFD da imaju štetne efekte na funkciju ovih proteina. Razlika od 7% je verovatno posledica dodatnih informacija koje PolyPhen-2 koristi i to, najznačajnije, parametra pfam-hit jer se ovaj parametar odnosi upravo na informaciju o poziciji analiziranog polimorfizma u proteinskim domenima definisanim u Pfam bazi CFD (Finn i sar. 2013). Dodatnu potvrdu značajne asociranosti predikcija PolyPhen-2 i SIFT sa CFD/nCFD pozicijom AAS dala je i Spirmanova korelaciona analiza.

Do sada nije bilo istraživanja koja su ispitivala efikasnost PolyPhen-2 i SIFT u predviđanju funkcionalnih efekata nCFD AAS, već su ispitivane performanse ovih alata nezavisno od pozicije AAS. Rezultati evaluacije na SVE_AAS skupu, koji sadrži sve AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2, su u skladu sa navedenim prethodnim istraživanjima. Najčešće korišćene baze AAS za evaluaciju performansi programa za predikciju funkcionalnih efekata AAS su HumDiv, koja sadrži mutacije asocirane sa Mendelovskim bolestima, i HumVar, koja sadrži sve mutacije iz UniProt, koje su definisane kao asocirane sa nekom bolešću, što uključuje i kompleksne bolesti. Na ovim test skupovima, PolyPhen-2 je postizao specifičnost od 80% i senzitivnost od 92% i 73% na HumDiv i HumVar, redom (Adzhubei i sar. 2010). SIFT je predviđao funkcionalne efekte AAS iz ovih baza sa senzitivnošću od 80% i specifičnošću od 87% i 75%, redom (Sim i sar. 2012). Thusberg i sar. su dobili slične, mada nešto niže vrednosti senzitivnosti i specifičnosti PolyPhen-2 i SIFT na bazi podataka, koja je sadržala 19 335 štetnih mutacija asociranih sa različitim humanim bolestima i 20 170 neutralnih SNP-ova (Thusberg i sar. 2011). Ovi autori su pokazali da je specifičnost PolyPhen-2 i SIFT u predviđanju funkcionalnih efekata AAS u njihovoj bazi 0.70 i 0.62, a senzitivnost 0.73 i 0.68, redom. Takođe, predikcije PolyPhen-2 i SIFT su korelirale na nivou 0.50, što je u skladu sa rezultatima dobijenim u ovom istraživanju ($r = 0.48$). Veća tačnost PolyPhen-2 na HumDiv i HumVar podacima u odnosu na SIFT i oba alata na podacima Thusberg i saradnika je posledica toga što je ovaj alat treniran na njima. Performansa PolyPhen-2 i SIFT na podacima Thusberg i saradnika je bliža performansi na HumVar nego na HumDiv podacima verovatno zbog prirode mutacija u bazama, ove dve baze sadrže i mutacije u kompleksnim bolestima, za razliku od HumDiv, koja sadrži samo mutacije u

Mendelovskim bolsetima. Izvršena je analiza performansi ovih alata i na 267 AAS iz četiri gena asocirana sa različitim vrstama raka – BRCA1, MSH2, MLH1 i TP53 (Hicks i sar. 2011). Specifičnost PolyPhen-2 i SIFT u predikciji ovih AAS je 0.54 i 0.46, a senzitivnost 0.85 i 0.83, redom. Rezultati dobijeni u ovoj disertaciji (na SVE_AAS skupu) i literaturni podaci ukazuju na to da je specifičnost PolyPhen-2 i SIFT u klasifikaciji AAS u kancer-asociranim genima relativno niska, dok je senzitivnost značajno veća.

Međutim, tačnost PolyPhen-2 i SIFT, a posebno element senzitivnosti, je značajno niža od očekivane ako se testiranje vrši na AAS pozicioniranim van CFD (nCFD_AAS). S obzirom da je pokazano da i mutacije u nCFD mogu da imaju važne funkcionalne efekte i da učestvuju u patogenezu bolesti (v. Poglavlje 1.3.4.), potrebno je razviti algoritam koji bi efikasno vršio ova predviđanja. Treba imati u vidu da su PolyPhen-2 i SIFT identifikovali mnoge od nCFD AAS kao štetne – PolyPhen-2 kategoriše 65 AAS, a SIFT 70 od 159 AAS u nCFD_AAS kao štetne mutacije. Međutim, samo 20, odnosno 26 AAS su tačno predviđene mutacije, dok su ostalo lažno pozitivni rezultati PolyPhen-2 i SIFT, redom. Dakle, evoluciona konzerviranost neke pozicije u sekvenci ne korelira sa njenim funkcionalnim značajem ukoliko ona nije deo CFD. Osnovna hipoteza u nastavku ovog dela istraživanja je da su funkcionalni efekti nCFD AAS posledica razlike u fizičko-hemijskim svojstvima supstitucijske aminokiseline u odnosu na divlji tip.

Važno je zapaziti da, iako PolyPhen-2 koristi neke osobine AAS koje ne zavise od evolucione konzerviranosti, tačnost njegovih predikcija na nCFD_AAS je manja u odnosu na SIFT. Ovo se može objasniti pretpostavkom da različiti parametri na kojima se zasniva PolyPhen-2, kao što je pfam-hit, doprinose upravo identifikaciji aminokiselinskih pozicija u sekvenci koje su deo CFD, što se ogleda i u značajnoj i visokoj korelaciji PolyPhen-2 predikcija i pozicije AAS u CFD/nCFD. S druge strane, predviđanja SIFT-a, koji ima nižu vrednost ovih korelacija, se isključivo odnose na konzerviranosti pojedinačnih pozicija u proteinskoj sekvenci, ne uzimajući u obzir da li je ta pozicija deo CFD ili ne.

CFD/nCFD distribucija mutacija u humanim bolestima zavisi od stope evolucionih promena samih mutiranih gena, koji su uključeni u patogenezu bolesti (Thomas PD i

Kejariwal 2004; Podder i Ghosh 2009). Evoluciona stopa gena uključenih u patogenezu poligenskih bolesti, kao što je kancer, je značajno viša u odnosu na gene uključene u monogenske, Mendelovske bolesti i određena je sa četiri faktora: i) prisustvom čvorova u PPI mrežama, ii) velikim brojem neuređenih aminokiselinskih rezidua, iii) nižim nivoom ekspresije i iv) uključenosti u regulatorne procese. 'Mendelovske' mutacije imaju izraženu tendenciju da se javljaju na visoko konzerviranim aminokiselinskim pozicijama u sekvencama (Miller i Kumar 2001; Ng i Henikoff 2002; Thomas PD i sar. 2003; Thomas PD i Kejariwal 2004). S druge strane, mutacije asocirane sa kompleksnim bolestima imaju različitu distribuciju, koja se ne razlikuje od distribucije 'normalnih' humanih varijacija i nije dominantno vezana za proteinske CFD (Thomas PD i Kejariwal 2004; Kumar i sar. 2011). Navedeni podaci i rezultati ove disertacije mogu objasniti i razliku u performansi PolyPhen-2 i SIFT na HumDiv i HumVar bazama. PolyPhen-2 i SIFT se uspešno koriste za identifikaciju mutacija asociranih sa Mendelovskim bolestima (npr. Dequeker i sar. 2008; Wang H i sar. 2010; Fan i sar. 2011; Ionita-Laza i sar. 2011; Sebro i sar. 2011; Masica i sar. 2012; Li J i sar. 2013), ali, takođe, postoji veliki broj radova u kojima se ovi alati, kao standard, koriste za predikciju funkcionalnih efekata AAS asociranih sa različitim kompleksnim bolestima, uključujući kancer (npr. Kaminker i sar. 2007; Desronvil i sar. 2010; Bogdanova i sar. 2011; Clague i sar. 2011; Hicks i sar. 2011; Landwehr i sar. 2011; Le Calvez-Kelm i sar. 2011; Lubbe i sar. 2011; Rossi i sar. 2011; Wong i sar. 2011; Liu J i sar. 2012; Pan i sar. 2012; Vermaat i sar. 2012; De Franco i sar. 2013; Shain i Pollack 2013). Rezultati dobijeni ovim istraživanjem i literaturni podaci ukazuju da bi fokus budućih modifikacija ovih alata, sa ciljem poboljšanja performanse predikcija mutacija asociranih sa kompleksnim bolestima, trebalo da bude na nCFD polimorfizmima, tj. poboljšanju diskriminacije između neutralnih SNP-ova i štetnih mutacija pozicioniranih u nCFD regionima proteina.

ISM algoritam vrši predviđanje funkcionalnih efekata AAS na osnovu promena značajnih frekventnih domena u informacionom spektru analiziranog protein (Gemovic i sar. 2013). S obzirom da se proteini uključeni u patogenezu kancera karakterišu velikim broj prolaznih (engl. *transient*) PPI (Jonsson i Bates 2006), utvrđivanje efekata AAS na

svojstva koja su uključena u ove PPI može doprineti tačnosti predviđanja funkcionalnih efekata AAS, kao i razumevanju patoloških efekata mutacija. EIIP ima ulogu u međumolekulskim interakcijama proteina, pa ISM algoritam procenjuje efekat AAS na značajne PPI ulaznog proteina tako što detektuje uticaj na distribuciju EIIP (promene u informacionom spektru). PPI-asocirana svojstva proteina su značajna za predikciju efekata nCFD AAS zbog toga što su nekonzervirane AAS često uključene u promiskuitetne, funkcionalno značajne interakcije proteina asociranih s kancerima (Mintseris i Weng 2005).

Ključni korak u ISM algoritmu za predikciju funkcionalnih efekata AAS je identifikacija značajne frekvence u informacionom spektru: i) koja najbolje razlikuje sekvence sa SNP-ovima od onih sa mutacijama i ii) povećanje/smanjenje amplitude, uzrokovano AAS, na ovoj frekvenci u odnosu na divlji tip ne zavisi od pozicije AAS u proteinskoj sekvenci (CFD/nCFD). Prvo svojstvo se detektuje statističkom analizom – neparametrijskim Mann-Whitney U testom. Drugo svojstvo se ispituje virtuelnim Ala skeniranjem. Za ASXL1, EZH2 i TET2 se pokazalo da su značajne frekvence $F(0,476)$, $F(0,411)$ i $F(0,491)$, redom. Mutacije u ovim proteinima povećavaju vrednosti amplitude na ovim frekvencama, dok ih SNP-ovi smanjuju. Za DNMT3A je kao značajna frekvencia identifikovana $F(0,071)$, pri čemu mutacije u ovom proteinu dovode do smanjenja amplitude na ovoj frekvenci, a SNP-ovi je povećavaju. Povećanje/smanjenje amplitude na značajnim frekvencama u informacionim spektrima sva četiri proteina ne zavisi od pozicije AAS u sekvencama ovih epigenetičkih regulatora.

Ovako identifikovani kriterijumi, sa graničnom vrednošću amplitude koja odgovara divljem tipu, su imali tačnost od ~70% u predikciji funkcionalnih efekata nCFD AAS, što je 17% i 12% bolje u odnosu na PolyPhen-2 i SIFT. Ovaj rezultat pokazuje da je pri analizi nCFD AAS bolje koristiti ISM algoritam za predviđanje njihovih funkcionalnih efekata. Takođe, ovi rezultati ukazuju na to da promene u distribuciji EIIP u sekvencama epigenetičkih regulatora sa mutacijama u odnosu na divlji tip zaista doprinose njihovom funkcionalnom efektu i uticaju na patogenezu mijeloidnih maligniteta.

U mijeloidnim malignitetima, postoje specifični profili metilacije DNK i modifikacija histona, pa važna grupa terapeutika obuhvata modulatore epigenetičkih modifikacija. Za kliničku upotrebu su odobreni inhibitori DNK-metiltransferaza i histon-deacetilaza (Silverman i sar. 2002; Issa i sar. 2004; Lyko i Brown 2005; Garcia-Manero i sar. 2008). Mutacije u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2 imaju važnu ulogu u genezi i progresiji svih tipova mijeloidnih maligniteta (pregled u Shih i sar. 2012). Ove mutacije predstavljaju i važne prognostičke biomarkere (Gelsi-Boyer i sar. 2010; Ley i sar. 2010; Bejar i sar. 2011 (a); Grossmann i sar. 2011; Lin i sar. 2011; Markova i sar. 2011; Metzeler i sar. 2011 (a), 2011 (b); Shen Y i sar. 2011; Thol i sar. 2011 (a), 2011 (b), 2011 (c); Walter i sar. 2011; Yan XJ i sar. 2011; Gelsi-Boyer i sar. 2012; Hajkova i sar. 2012; Hou i sar. 2012; Patel i sar. 2012; Pratcorona i sar. 2012; Renneville i sar. 2012; Schnittger i sar. 2013; Patnaik i sar. 2014). Mutacije u DNMT3A i TET2 imaju važnu ulogu u odgovoru pacijenata na terapiju, pre svega, na inhibitore DNK-metiltransferaza, kao što je azacitidin (Itzykson i sar. 2011; Sekeres i sar. 2012), pa se mogu koristiti kao biomarkeri za prateću dijagnostiku.

Da bi se mutacije u epigenetičkim regulatorima koristile kao klinički prognostički i biomarkeri pri terapijskom odlučivanju, neophodno ih je precizno razdvojiti od neutralnih SNP-ova (Tchernitchko i sar. 2004; Lee TC i sar. 2008; Crockett i sar. 2012). Eksperimentalno testiranje somatskog statusa i patogenog efekta AAS je skupo i vremenski zahtevno, pa je važno razvijati kompjuterske metode. Oko 74% štetnih mutacija u ovim proteinima je pozicionirana u njihovim CFD, pa ih MSA-zasnovani alati, kao što su PolyPhen-2 i SIFT, dobro detektuju. Međutim, više od četvrtine štetnih mutacija u ovim proteinima su u nCFD, pa je, s obzirom da ih PolyPhen-2 i SIFT ne razlikuju efikasno od neutralnih SNP-ova, neophodno koristiti nove algoritme, kao što je ISM. Međutim, tačnost od ~70% u predviđanju funkcionalnih efekata nCFD AAS, nije dovoljno visoka za kliničku upotrebu i zahteva dalje unapređenje, koje se može postići na nekoliko načina.

Ako uzmemo u obzir pretpostavku da su različiti lokalni maksimumi u informacionom spektru proteina asocirani sa različitim biološkim funkcijama i interakcijama tog proteina, unapređenje ISM algoritma bi trebalo zasnivati na kreiranju ISM kriterijuma koji bi uzimao

u obzir više od jedne frekvence. Ovo bi posebno bilo značajno za proteine koji su mutirani u različitim bolestima, sa različitim patofiziološkim osnovama. Primer korisnosti ovakve strategije je EZH2, u kome je identifikovano šest nCFD mutacija, pri čemu su tri tačno klasifikovane ISM algoritmom (L149Q, A384T i T568I), dok su druge tri lažno negativno anotirane kao SNP-ovi (M134K, C534R i L575P). Detaljno ispitivanje ovih predikcija je pokazalo da su sve tačno klasifikovane mutacije detektovane kod pacijenata sa MPN, dok su sve netačno klasifikovane mutacije poreklom iz pacijenata sa MDS. Ovaj rezultat ukazuje na to da promene na ISM frekvenci $F(0,411)$ možda koreliraju sa disfunkcijom u proliferaciji koja vodi ka MPN, a ne u diferencijaciji, što je u osnovi MDS (Vardiman 2010).

Dalji rad na poboljšanju kriterijuma za anotaciju nCFD AAS bi trebalo da obuhvati i druge fizičko-hemijske karakteristike aminokiselina, koje utiču na funkcionisanje proteina. S obzirom da EIIP pripada grupi parametara koja se odnosi na elektrostatička i elektrodinamička svojstva aminokiselina, dodatnu informaciju za analizu funkcionalnih efekata AAS bi pružilo uključivanje svojstava iz drugih grupa (Atchley i sar. 2005). Na taj način bi se kombinacijom neredundantnih osobina potencijalno povećala moć predikcionog algoritma. Hidrofobnost je osobina čiji se strukturni obrasci, kao i u sličaju EIIP, detektuju Furijeovom transformacijom (Leonov i Arkin 2005; Guo YZ i sar. 2006; Illingworth i sar. 2008; Marsella i sar. 2009). Ona se široko primenjuje za detekciju proteinskih regiona uključenih u *foldings* (Eisenberg i sar. 1984; Cornette i sar. 1987; Xiong i sar. 1995; Murray i sar. 2002; Sepher i sar. 2003; Shen HB i Chou 2006; Marsella i sar. 2009; Zhang S i Wang 2009; Meher i sar. 2012), pa se može pretpostaviti da bi se uvođenjem ovog parametra u ISM algoritam mogle detektovati i mutacije koje utiču na sekundarnu i tercijsku strukturu proteina.

Otkriće i razvoj lekova je kompleksan i skup proces. Savremene tendencije uključuju nekoliko prekliničkih faza ispitivanja, a bioinformatika ima važnu ulogu u prvim koracima ovog procesa. Najznačajnija primena bioinformatičkih metoda je u identifikaciji targeta i dizajnu i optimizaciji jedinjenja, koja su kandidati za lek. Važan dodatak ciljanoj terapiji, za postizanje personalizovanog pristupa, je prateća dijagnostika, koja omogućava selekciju

pacijenata, na osnovu biomarkera, koji će imati benefit od terapije. Zbog velike heterogenosti maligniteta, prateća dijagnostika ima veliki značaj u onkologiji. Najčešći informativni biomarkeri su mutacije u kancerskim genima. Ovo istraživanje, sprovedeno na ključnim faktorima uključenim u patogenezu mijeloidnih maligniteta, je pokazalo da je ISM efikasan za identifikaciju molekula i interakcija, koji su terapijski targeti. Takođe, ISM predstavlja komplementarnu metodu postojećim programima za detekciju funkcionalnih mutacija, koje su dijagnostički i prognostički biomarkeri mijeloidnih maligniteta.

6. Zaključci

Cilj ovog istraživanja je obuhvatao predviđanje novih PPI proteina WT1, identifikaciju i karakterizaciju funkcionalnih domena proteina NISCH i razvoj algoritma za predviđanje funkcionalnih efekata AAS u epigenetičkim regulatorima ASXL1, DNMT3A, EZH2 i TET2, primenom bioinformatičkih metoda. Dobijeni rezultati ukazuju na to da je ISM efikasan za identifikaciju molekula, molekulskih domena i PPI, koji su kandidati za terapijske targete. Posebno je značajno da je ISM-zasnovani algoritam, koji je prvi put primenjen specifično na analizu funkcionalnih efekata mutacija van konzerviranih domena proteina, efikasan u stratifikaciji mutacija i SNP-ova i predstavlja komplementarnu metodu postojećim programima PolyPhen-2 i SIFT.

Iz celokupnog ljudskog proteoma, izdvojeni su, pomoću integrativnog algoritma za predikciju PPI, kandidati za nove interaktore WT1 sa sledećim karakteristikama: i) identične determinante informacionog spektra sa WT1; ii) učestvuju u regulaciju transkripcije; iii) u najbližem su susedstvu WT1 u mreži humanih PPI. Pokazano je da je dominantna frekvencija informacionog spektra WT1 prisutna u informacionim spektrima proteina uključenih u sve osnovne mehanizme regulacije ekspresije gena – transkripciju, iRNK splajsovanje i epigenetičke modifikacije. Ovaj rezultat može da objasni mehanizam kojim WT1 učestvuje u regulaciji 'iRNK fabrika'.

Ekspertskom analizom kandidata koekspimiranih sa WT1 u K562 ćelijskoj liniji, utvrđeno je da je potencijalna PPI koja je najrelevantniji kandidat za ciljanu terapiju, interakcija WT1 i ciklin zavisne kinase, CDK9. Ova PPI je, nakon predikcije u okviru ove disertacije, eksperimentalno i potvrđena u laboratoriji sa kojom Institut za nuklearne nauke Vinča ima međunarodnu saradnju. Ovaj rezultat se može smatrati potvrdom efikasnosti integrativnog algoritma za predikciju PPI.

Bioinformatičkom analizom proteinske sekvence tumor supresora NISCH-a, koji je i kandidat za I₁ imidazolinski receptor, identifikovana su dva funkcionalna domena: (1) 680-

739ak i (2) 966-1020ak. Prvi domen je funkcionalno anotiran procesima uključenim u: i) sinaptogenezu i ii) post-translacione modifikacije. Ovi rezultati su u skladu sa literaturnim podacima. Drugi domen NISCH-a je prvi put u ovoj disertaciji funkcionalno anotiran. Pokazano je da je uključen u: i) apoptozu, ii) transkripciju i iii) signalnu transdukciju posredovanu hormonima. Ovaj rezultat ima primenu u identifikaciji malih molekula koji reaguju sa NISCH-om i utiču na biološke procese. Nakon što je pokazano da rilmenidin, agonist I₁-IR, može da izazove apoptozu u leukemijskoj ćelijskoj liniji, naknadno 3D modelovanje je pokazalo da su mesta predviđena ISM-om najverovatnije mesta vezivanja rilmenidina, ali ovaj rezultat čeka eksperimentalnu potvrdu.

U okviru ovog istraživanja je razvijen ISM-zasnovani algoritam za predviđanje funkcionalnih efekata AAS. Njegova efikasnost je upoređena sa najčešće korišćenim programima, PolyPhen-2 i SIFT, na skupu AAS u nekonzerviranim regionima epigenetičkih regulatora ASXL1, EZH2, DNMT3A i TET2. Pokazano je da jedino ISM algoritam statistički značajno diferencira štetne mutacije od neutralnih SNP-ova u analiziranom skupu, sa tačnošću ~70%, što je 17% i 12% više u odnosu na PolyPhen-2 i SIFT, redom. U skladu sa dobijenim rezultatima, preporučena je upotreba analize zasnovane na postulatima informacionih spektara, kao komplementarne MSA i strukturnoj.

7. Literatura

1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 491(7422): 56-65.

Abbas S, Lugthart S, Kavelaars FG, Schelen A, Koenders JE, Zeilemaker A, van Putten WJ, Rijnveld AW, Lowenberg B, Valk PJ. 2010. Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood*. 116(12): 2122-2126.

Abdel-Wahab O i sar. 2012. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell*. 22(2): 180-193.

Aceros H, Farah G, Cobos-Puc L, Stabile AM, Noiseux N, Mukaddam-Daher S. 2011. Moxonidine improves cardiac structure and performance in SHR through inhibition of cytokines, p38 MAPK and Akt. *Br J Pharmacol*. 164(3): 946-957.

Aceros H, Farah G, Noiseux N, Mukaddam-Daher S. 2014. Moxonidine modulates cytokine signalling and effects on cardiac cell viability. *Eur J Pharmacol*. 740: 168-182.

Adamcsek B, Palla G, Farkas IJ, Derényi I, Vicsek T. 2006. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*. 22(8): 1021-1023.

Adams JM, Cory S. 1998. The Bcl-2 protein family: arbiters of cell survival. *Science*. 281(5381): 1322-1326.

Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. Chapter 7: Unit7.20.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods*. 7(4): 248-249.

Alahari SK. 2003. Nischarin inhibits Rac induced migration and invasion of epithelial cells by affecting signaling cascades involving PAK. *Exp Cell Res*. 288(2): 415-424.

Alahari SK, Lee JW, Juliano RL. 2000. Nischarin, a novel protein that interacts with the integrin alpha5 subunit and inhibits cell migration. *J Cell Biol*. 151(6): 1141-1154.

Alahari SK, Reddig PJ, Juliano RL. 2004. The integrin-binding protein Nischarin regulates cell migration by inhibiting PAK. *EMBO J*. 23(14): 2777-2788.

Algar EM, Khromykh T, Smith SI, Blackburn DM, Bryson GJ, Smith PJ. 1996. A WT1 antisense oligonucleotide inhibits proliferation and induces apoptosis in myeloid leukaemia cell lines. *Oncogene*. 12(5): 1005-1014.

Al-Lazikani B, Banerji U, Workman P. 2012. Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol.* 30(7): 679-692.

Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, Kanaya S. 2006. Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics.* 7: 207.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17): 3389-3402.

Alvarez S i sar. 2010. DNA methylation profiles and their relationship with cytogenetic status in adult acute myeloid leukemia. *PLoS One.* 5(8): e12197.

Appelbaum FR, Gundacker H, Head DR, Slovak ML, Willman CL, Godwin JE, Anderson JE, Petersdorf SH. 2006. Age and acute myeloid leukemia. *Blood.* 107(9): 3481-3485.

Aravind L, Iyer LM. 2012. The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle.* 11(1): 119-131.

Armstrong JF, Pritchard-Jones K, Bickmore WA, Hastie ND, Bard JB. 1993. The expression of the Wilms' tumour gene, WT1, in the developing mammalian embryo. *Mech Dev.* 40(1-2): 85-97.

Asim M, Hafeez BB, Siddiqui IA, Gerlach C, Patz M, Mukhtar H, Baniahmad A. 2011. Ligand-dependent corepressor acts as a novel androgen receptor corepressor, inhibits prostate cancer growth, and is functionally inactivated by the Src protein kinase. *J Biol Chem.* 286(43): 37108-37117.

Atchley WR, Zhao J, Fernandes AD, Drüke T. 2005. Solving the protein sequence metric problem. *Proc Natl Acad Sci USA.* 102(18): 6395-6400.

Bader GD, Hogue CW. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 4: 2.

Bair CL, Oppenheim A, Trostel A, Prag G, Adhya S. 2008. A phage display system designed to detect and study protein-protein interactions. *Mol Microbiol.* 67(4): 719-728.

Bals J, Meyer J, Mueller W, Korshunov A, Hartmann C, von Deimling A. 2008. Analysis of the IDH1 codon 132 mutation in brain tumors. *Acta Neuropathol.* 116(6): 597-602.

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res.* 21(3): 381-395.

Bansal H i sar. 2010. Heat shock protein 90 regulates the expression of Wilms tumor 1 protein in myeloid leukemias. *Blood.* 116(22): 4591-4599.

Bao L, Zhou M, Cui Y. 2005. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* 33(Web Server issue): W480-W482.

Baranwal S, Wang Y, Rathinam R, Lee J, Jin L, McGoey R, Pylayeva Y, Giancotti F, Blobel GC, Alahari SK. 2011. Molecular characterization of the tumor-suppressive function of nischarin in breast cancer. *J Natl Cancer Inst.* 103(20): 1513-1528.

Barboric M, Nissen RM, Kanazawa S, Jabrane-Ferrat N, Peterlin BM. 2001. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol Cell.* 8(2): 327-337.

Bardeesy N, Pelletier J. 1998. Overlapping RNA and DNA binding domains of the wt1 tumor suppressor gene product. *Nucleic Acids Res.* 26(7): 1784-1792.

Barragán E, Cervera J, Bolufer P, Ballester S, Martín G, Fernández P, Collado R, Sayas MJ, Sanz MA. 2004. Prognostic implications of Wilms' tumor gene (WT1) expression in patients with de novo acute myeloid leukemia. *Haematologica.* 89(8): 926-933.

Barrett T et al. 2013. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41(Database issue): D991-D995.

Baskind HA, Na L, Ma Q, Patel MP, Geenen DL, Wang QT. 2009. Functional conservation of Asxl2, a murine homolog for the Drosophila enhancer of trithorax and polycomb group gene Asx. *PLoS One.* 4(3): e4750.

Baumli S, Endicott JA, Johnson LN. 2010. Halogen bonds form the basis for selective P-TEFb inhibition by DRB. *Chem Biol.* 17(9): 931-936.

Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, Debreczeni JE, Knapp S, Johnson LN. 2008. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J.* 27(13): 1907-1918.

Baylin SB. 2005. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol.* 2 Suppl 1: S4-S11.

Bejar R et al. 2011 (a). Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med.* 364(26): 2496-2506.

Bejar R, Levine R, Ebert BL. 2011 (b). Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J Clin Oncol.* 29(5): 504-515.

Ben-Hur A, Noble WS. 2005. Kernel methods for predicting protein-protein interactions. *Bioinformatics.* 21 Suppl 1: I38-46.

Bentley D. 2002. The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol.* 14(3): 336-342.

Bentley DL. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol.* 17(3): 251-256.

Bestor TH. 2000. The DNA methyltransferases of mammals. *Hum Mol Genet.* 9(16): 2395-2402.

Bhattacharya S, Eckner R, Grossman S, Oldread E, Arany Z, D'Andrea A, Livingston DM. 1996. Cooperation of Stat2 and p300/CBP in signalling induced by interferon-alpha. *Nature.* 383(6598): 344-347.

Bid HK, Roberts RD, Manchanda PK, Houghton PJ. 2013. RAC1: an emerging therapeutic option for targeting cancer angiogenesis and metastasis. *Mol Cancer Ther.* 12(10): 1925-1934.

Bielli P, Busà R, Paronetto MP, Sette C. 2011. The RNA-binding protein Sam68 is a multifunctional player in human cancer. *Endocr Relat Cancer.* 18(4): R91-R102.

Bienengraeber M i sar. 2004. ABCC9 mutations identified in human dilated cardiomyopathy disrupt catalytic KATP channel gating. *Nat Genet.* 36(4): 382-387.

Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics.* 25(22): 3045-3046.

Bock JR, Gough DA. 2001. Predicting protein--protein interactions from primary structure. *Bioinformatics.* 17(5): 455-460.

Bogdanova N, Sokolenko AP, Iyevleva AG, Aбыsheva SN, Blaut M, Bremer M, Christiansen H, Rave-Fränk M, Dörk T, Imyanitov EN. 2011. PALB2 mutations in German and Russian patients with bilateral breast cancer. *Breast Cancer Res Treat.* 126(2): 545-550.

Bor YC, Swartz J, Morrison A, Rekosh D, Lodomery M, Hammarskjöld ML. 2006. The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. *Genes Dev.* 20(12): 1597-1608.

Bossi A, Lehner B. 2009. Tissue specificity and the human protein interaction network. *Mol Syst Biol.* 5: 260.

Boultonwood J, Wainscoat JS. 2007. Gene silencing by DNA methylation in haematological malignancies. *Br J Haematol.* 138(1): 3-11.

Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. 2003. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* 22(20): 5323-5335.

Bradley AP. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* 30.7: 1145-1159.

Brain BJ. 2010. Myeloid and lymphoid neoplasms with eosinophilia and abnormalities of PDGFRA, PDGFRB or FGFR1. *Haematologica*. 95(5): 696-698.

Bromberg Y, Rost B. 2007. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*. 35(11): 3823-3835.

Brown CJ i sar. 2013. Stapled peptides with improved potency and specificity that activate p53. *ACS Chem Biol*. 8(3): 506-512.

Brown NR, Noble ME, Endicott JA, Johnson LN. 1999. The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nat Cell Biol*. 1(7): 438-443.

Byrd JC i sar.; Cancer and Leukemia Group B (CALGB 8461). 2002. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood*. 100(13): 4325-4336.

Cai D, Latham VM Jr, Zhang X, Shapiro GI. 2006. Combined depletion of cell cycle and transcriptional cyclin-dependent kinase activities induces apoptosis in cancer cells. *Cancer Res*. 66(18): 9270-9280.

Call KM i sar. 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell*. 60(3): 509-520.

Cancelas JA, Williams DA. 2009. Rho GTPases in hematopoietic stem cell functions. *Curr Opin Hematol*. 16(4): 249-254.

Cao R, Zhang Y. 2004. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell*. 15(1): 57-67.

Capriotti E, Calabrese R, Casadio R. 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 22(22): 2729-2734.

Carbonell P, Nussinov R, del Sol A. 2009. Energetic determinants of protein binding specificity: insights into protein interaction networks. *Proteomics*. 9(7): 1744-1753.

Caricasole A, Duarte A, Larsson SH, Hastie ND, Little M, Holmes G, Todorov I, Ward A. 1996. RNA binding by the Wilms tumor suppressor zinc finger proteins. *Proc Natl Acad Sci USA*. 93(15): 7562-7566.

Carpenter B i sar. 2004. BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1. *Mol Cell Biol*. 24(2): 537-549.

Chan SS, Longley MJ, Copeland WC. 2005. The common A467T mutation in the human mitochondrial DNA polymerase (POLG) compromises catalytic efficiency and interaction with the accessory subunit. *J Biol Chem*. 280(36): 31341-31346.

Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S. 2009. MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics*. 25(5): 690-691.

Cheema SK, Mishra SK, Rangnekar VM, Tari AM, Kumar R, Lopez-Berestein G. 2003. Par-4 transcriptionally regulates Bcl-2 through a WT1-binding site on the bcl-2 promoter. *J Biol Chem*. 278(22): 19995-20005.

Chou KC. 2000. Prediction of protein subcellular locations by incorporating quasi-sequence-order effect. *Biochem Biophys Res Commun*. 278(2): 477-483.

Chou KC. 2001. Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins*. 43(3): 246-255.

Chou KC. 2005. Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. *Bioinformatics*. 21(1): 10-19.

Chou KC. 2011. Some remarks on protein attribute prediction and pseudo amino acid composition. *J Theor Biol*. 273(1): 236-247.

Chou KC, Cai YD. 2005. Prediction of membrane protein types by incorporating amphipathic effects. *J Chem Inf Model*. 45(2): 407-413.

Christopoulos A. 2002. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat Rev Drug Discov*. 1(3): 198-210.

Ciechanover A. 1994. The ubiquitin-proteasome proteolytic pathway. *Cell*. 79(1): 13-21.

Cilloni D i sar. 2002. Quantitative assessment of WT1 expression by real time quantitative PCR may be a useful tool for monitoring minimal residual disease in acute leukemia patients. *Leukemia*. 16(10): 2115-2121.

Cilloni D i sar. 2009. Real-time quantitative polymerase chain reaction detection of minimal residual disease by standardized WT1 assay to enhance risk stratification in acute myeloid leukemia: a European LeukemiaNet study. *J Clin Oncol*. 27(31): 5195-5201.

Clague J, Wilhoite G, Adamson A, Bailis A, Weitzel JN, Neuhausen SL. 2011. RAD51C germline mutations in breast and ovarian cancer cases from high-risk families. *PLoS One*. 6(9): e25632.

Conaway RC, Conaway JW. 2011. Function and regulation of the Mediator complex. *Curr Opin Genet Dev*. 21(2): 225-230.

Cools J, Mentens N, Furet P, Fabbro D, Clark JJ, Griffin JD, Marynen P, Gilliland DG. 2004. Prediction of resistance to small molecule FLT3 inhibitors: implications for molecularly targeted therapy of acute leukemia. *Cancer Res*. 64(18): 6385-6389.

Cornette JL, Cease KB, Margalit H, Spouge JL, Berzofsky JA, DeLisi C. 1987. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J Mol Biol.* 195(3): 659-685.

Cosic I, Drummond AE, Underwood JR, Hearn MT. 1994. In vitro inhibition of the actions of basic FGF by a novel 16 amino acid peptide. *Mol Cell Biochem.* 130(1): 1-9.

Cosic I, Hearn MT. 1991. 'Hot spot' amino acid distribution in Ha-ras oncogene product p21: relationship to guanine binding site. *J Mol Recognit.* 4(2-3): 57-62.

Ćosić I, Nesić D, Pavlović M, Williams R. 1986. Enhancer binding proteins predicted by informational spectrum method. *Biochem Biophys Res Commun.* 141(2): 831-838.

Cosic I, Pavlovic M, Vojisavljevic V. 1989. Prediction of "hot spots" in interleukin-2 based on informational spectrum characteristics of growth-regulating factors. Comparison with experimental data. *Biochimie.* 71(3): 333-342.

Couronné L, Bastard C, Bernard OA. 2012. TET2 and DNMT3A mutations in human T-cell lymphoma. *N Engl J Med.* 366(1): 95-96.

Crockett DK, Lyon E, Williams MS, Narus SP, Facelli JC, Mitchell JA. 2012. Utility of gene-specific algorithms for predicting pathogenicity of uncertain gene variants. *J Am Med Inform Assoc.* 19(2): 207-211.

Dalva MB, Takasu MA, Lin MZ, Shamah SM, Hu L, Gale NW, Greenberg ME. 2000. EphB receptors interact with NMDA receptors and regulate excitatory synapse formation. *Cell.* 103(6): 945-956.

Dameshek W. 1951. Some speculations on the myeloproliferative syndromes. *Blood.* 6(4): 372-375.

Dang L i sar. 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature.* 462(7274): 739-744.

Darie CC, Deinhardt K, Zhang G, Cardasis HS, Chao MV, Neubert TA. 2011. Identifying transient protein-protein interactions in EphB2 signaling by blue native PAGE and mass spectrometry. *Proteomics.* 11(23): 4514-4528.

Davies RC, Calvio C, Bratt E, Larsson SH, Lamond AI, Hastie ND. 1998. WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes Dev.* 12(20): 3217-3225.

De Franco E, Shaw-Smith C, Flanagan SE, Shepherd MH; International NDM Consortium, Hattersley AT, Ellard S. 2013. GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes.* 62(3): 993-997.

de Trad CH, Fang Q, Cosic I. 2000. The resonant recognition model (RRM) predicts amino acid residues in highly conserved regions of the hormone prolactin (PRL). *Biophys Chem.* 84(2): 149-157.

Dequeker E i sar. 2009. Best practice guidelines for molecular genetic diagnosis of cystic fibrosis and CFTR-related disorders--updated European recommendations. *Eur J Hum Genet.* 17(1): 51-65.

Depping R, Schindler SG, Jacobi C, Kirschner KM, Scholz H. 2012. Nuclear transport of Wilms' tumour protein Wt1 involves importins α and β . *Cell Physiol Biochem.* 29(1-2): 223-232.

Desronvil T i sar. 2010. Distribution of COL8A2 and COL8A1 gene variants in Caucasian primary open angle glaucoma patients with thin central corneal thickness. *Mol Vis.* 16: 2185-2191.

Ding Y, Milosavljevic T, Alahari SK. 2008. Nischarin inhibits LIM kinase to regulate cofilin phosphorylation and cell invasion. *Mol Cell Biol.* 28(11): 3742-3756.

Ding Y, Zhang R, Zhang K, Lv X, Chen Y, Li A, Wang L, Zhang X, Xia Q. 2013. Nischarin is differentially expressed in rat brain and regulates neuronal migration. *PLoS One.* 8(1): e54563.

Disenza MT, He S, Lee TH, Chu LL, Bolon B, Goodyer P, Eccles M, Pelletier J. 2003. WT1 is a modifier of the Pax2 mutant phenotype: cooperation and interaction between WT1 and Pax2. *Oncogene.* 22(50): 8145-8155.

Döhner H i sar.; European LeukemiaNet. 2010. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood.* 115(3): 453-474.

Doliana R, Veljkovic V, Prljic J, Veljkovic N, De Lorenzo E, Mongiat M, Ligresti G, Marastoni S, Colombatti A. 2008. EMILINs interact with anthrax protective antigen and inhibit toxin action in vitro. *Matrix Biol.* 27(2): 96-106.

Dontenwill M i sar. 2003 (a). IRAS, the human homologue of Nischarin, prolongs survival of transfected PC12 cells. *Cell Death Differ.* 10(8): 933-935.

Druker BJ i sar. 2001. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med.* 344(14): 1031-1037.

Du X, Hublitz P, Günther T, Wilhelm D, Englert C, Schüle R. 2002. The LIM-only coactivator FHL2 modulates WT1 transcriptional activity during gonadal differentiation. *Biochim Biophys Acta.* 1577(1): 93-101.

Duffy MJ, Crown J. 2013. Companion biomarkers: paving the pathway to personalized treatment for cancer. *Clin Chem.* 59(10): 1447-1456.

Eberhardy SR, Farnham PJ. 2001. c-Myc mediates activation of the cad promoter via a post-RNA polymerase II recruitment mechanism. *J Biol Chem.* 276(51): 48562-48571.

Eberhardy SR, Farnham PJ. 2002. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J Biol Chem.* 277(42): 40156-40162.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30(1): 207-210.

Edgar RC, Batzoglou S. 2006 Multiple sequence alignment. *Curr Opin Struct Biol.* 16(3): 368-373.

Edmondson DG, Lyons GE, Martin JF, Olson EN. 1994. Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development.* 120(5): 1251-1263.

Edwards L, Ernsberger P. 2003. The I(1)-imidazoline receptor in PC12 pheochromocytoma cells reverses NGF-induced ERK activation and induces MKP-2 phosphatase. *Brain Res.* 980(1): 71-79.

Ehrlich M. 2009. DNA hypomethylation in cancer cells. *Epigenomics.* 1(2): 239-259.

Eisenberg D, Weiss RM, Terwilliger TC. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA.* 81(1): 140-144.

Ellis JD i sar. 2012. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell.* 46(6): 884-892.

Ellisen LW, Carlesso N, Cheng T, Scadden DT, Haber DA. 2001. The Wilms tumor suppressor WT1 directs stage-specific quiescence and differentiation of human hematopoietic progenitor cells. *EMBO J.* 20(8): 1897-1909.

Englert C, Maheswaran S, Garvin AJ, Kreidberg J, Haber DA. 1997. Induction of p21 by the Wilms' tumor suppressor gene WT1. *Cancer Res.* 57(8): 1429-1434.

Ernst T i sar. 2010. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet.* 42(8): 722-726.

Essafi A i sar. 2011. A wt1-controlled chromatin switching mechanism underpins tissue-specific wnt4 activation and repression. *Dev Cell.* 21(3): 559-574.

Estey E, Döhner H. 2006. Acute myeloid leukaemia. *Lancet.* 368(9550): 1894-1907.

Eswaran J, Soundararajan M, Kumar R, Knapp S. 2008. UnPAKING the class differences among p21-activated kinases. *Trends Biochem Sci.* 33(8): 394-403.

Ezhkova E, Pasolli HA, Parker JS, Stokes N, Su IH, Hannon G, Tarakhovskiy A, Fuchs E. 2009. Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell.* 136(6): 1122-1135.

Fan JY, Han B, Qiao J, Liu BL, Ji YR, Ge SF, Song HD, Fan XQ. 2011. Functional study on a novel missense mutation of the transcription factor FOXL2 causes blepharophimosis-ptosis-epicanthus inversus syndrome (BPES). *Mutagenesis*. 26(2): 283-289.

Fathi AT, Abdel-Wahab O. 2012. Mutations in epigenetic modifiers in myeloid malignancies and the prospect of novel epigenetic-targeted therapy. *Adv Hematol*. 2012: 469592.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Dostupno na: <http://globocan.iarc.fr>

Fernandez-Mercado M i sar. 2012. Mutation patterns of 16 genes in primary and secondary acute myeloid leukemia (AML) with normal cytogenetics. *PLoS One*. 7(8): e42334.

Ferrara N, Gerber HP, LeCouter J. 2003. The biology of VEGF and its receptors. *Nat Med*. 9(6): 669-676.

Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. 2005. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*. 21(14): 3176-3178.

Figueroa ME i sar. 2010 (a). DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell*. 17(1): 13-27.

Figueroa ME i sar. 2010 (b). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 18(6): 553-567.

Finn RD i sar. 2014. Pfam: the protein families database. *Nucleic Acids Res*. 42(Database issue): D222-D230.

Florio F, Cesaro E, Montano G, Izzo P, Miles C, Costanzo P. 2010. Biochemical and functional interaction between ZNF224 and ZNF255, two members of the Kruppel-like zinc-finger protein family and WT1 protein isoforms. *Hum Mol Genet*. 19(18): 3544-3556.

Foote J, Eisen HN. 1995. Kinetic and affinity limits on antibodies produced during immune responses. *Proc Natl Acad Sci USA*. 92(5): 1254-1256.

Forbes SA i sar. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*. 39(Database issue): D945-D950.

Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. 2008. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet*. Chapter 10: Unit 10.11.

Fried I, Bodner C, Pichler MM, Lind K, Beham-Schmid C, Quehenberger F, Sperr WR, Linkesch W, Sill H, Wölfler A. 2012. Frequency, onset and clinical impact of somatic DNMT3A mutations in therapy-related and secondary acute myeloid leukemia. *Haematologica*. 97(2): 246-250.

Fröhlich H. 1968. Long-range coherence and energy storage in biological systems. *Int J Quantum Chem*. 2(5): 641-649.

Fröhlich H. 1970. Long range coherence and the action of enzymes. *Nature*. 228(5276): 1093.

Fröhlich H. 1975. The extraordinary dielectric properties of biological materials and the action of enzymes. *Proc Natl Acad Sci USA*. 72(11): 4211-4215.

Fu TJ, Peng J, Lee G, Price DH, Flores O. 1999. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. *J Biol Chem*. 274(49): 34527-34530.

Fujii M i sar. 2006. SNIP1 is a candidate modifier of the transcriptional activity of c-Myc on E box-dependent target genes. *Mol Cell*. 24(5): 771-783.

Fujita T, Reis LF, Watanabe N, Kimura Y, Taniguchi T, Vilcek J. 1989. Induction of the transcription factor IRF-1 and interferon-beta mRNAs by cytokines and activators of second-messenger pathways. *Proc Natl Acad Sci U S A*. 86(24): 9936-9940.

Gaidzik VI i sar.; German-Austrian AML Study Group. 2009. Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood*. 113(19): 4505-4511.

Galm O, Herman JG, Baylin SB. 2006. The fundamental role of epigenetics in hematopoietic malignancies. *Blood Rev*. 20(1): 1-13.

Gao L, Bellantuono I, Elsässer A, Marley SB, Gordon MY, Goldman JM, Stauss HJ. 2000. Selective elimination of leukemic CD34(+) progenitor cells by cytotoxic T lymphocytes specific for WT1. *Blood*. 95(7): 2198-2203.

Gao SM, Xing CY, Chen CQ, Lin SS, Dong PH, Yu FJ. 2011. miR-15a and miR-16-1 inhibit the proliferation of leukemic cells by down-regulating WT1 protein level. *J Exp Clin Cancer Res*. 30: 110.

Garcia-Manero G i sar. 2008. Phase 1 study of the histone deacetylase inhibitor vorinostat (suberoylanilide hydroxamic acid [SAHA]) in patients with advanced leukemias and myelodysplastic syndromes. *Blood*. 111(3): 1060-1066.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 20(3): 307-315.

Gelsi-Boyer V i sar. 2010. ASXL1 mutation is associated with poor prognosis and acute transformation in chronic myelomonocytic leukaemia. *Br J Haematol.* 151(4): 365-375.

Gelsi-Boyer V, Brecqueville M, Devillier R, Murati A, Mozziconacci MJ, Birnbaum D. 2012. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *J Hematol Oncol.* 5: 12.

Gemovic B, Perovic V, Glisic S, Veljkovic N. 2013. Feature-based classification of amino acid substitutions outside conserved functional protein domains. *ScientificWorldJournal.* 2013: 948617.

Gene Ontology Consortium. 2001. Creating the gene ontology resource: design and implementation. *Genome Res.* 11(8): 1425-1433.

Gene Ontology Consortium [Internet]. c.1999-2014. Dostupno na: <http://geneontology.org>

Gentleman RC i sar. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5(10): R80.

Gessler M, Poustka A, Cavenee W, Neve RL, Orkin SH, Bruns GA. 1990. Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature.* 343(6260): 774-778.

Ghobrial IM, Witzig TE, Adjei AA. 2005. Targeting apoptosis pathways in cancer therapy. *CA Cancer J Clin.* 55(3): 178-194.

Gilliland DG. 2001. Hematologic malignancies. *Curr Opin Hematol.* 8(4): 189-191.

Glisic S, Arrigo P, Alavantic D, Perovic V, Prljic J, Veljkovic N. 2008. Lipoprotein lipase: A bioinformatics criterion for assessment of mutations as a risk factor for cardiovascular disease. *Proteins.* 70(3): 855-862.

Gojo I, Zhang B, Fenton RG. 2002. The cyclin-dependent kinase inhibitor flavopiridol induces apoptosis in multiple myeloma cells through transcriptional repression and down-regulation of Mcl-1. *Clin Cancer Res.* 8(11): 3527-3538.

Gonzalez de Castro D, Clarke PA, Al-Lazikani B, Workman P. 2013. Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance. *Clin Pharmacol Ther.* 93(3): 252-259.

Gordon S, Akopyan G, Garban H, Bonavida B. 2006. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene.* 25(8): 1125-1142.

Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, Jarrard DF, Isaacs WB, Pitha PM, Davidson NE, Baylin SB. 1995. E-cadherin expression is silenced by DNA

hypermethylation in human breast and prostate carcinomas. *Cancer Res.* 55(22): 5195-5199.

Green LM, Wagner KJ, Campbell HA, Addison K, Roberts SG. 2009. Dynamic interaction between WT1 and BASP1 in transcriptional regulation during differentiation. *Nucleic Acids Res.* 37(2): 431-440.

Greenman C i sar. 2007. Patterns of somatic mutation in human cancer genomes. *Nature.* 446(7132): 153-158.

Griffiths EA, Gore SD, Hooker CM, Mohammad HP, McDevitt MA, Smith BD, Karp JE, Herman JG, Carraway HE. 2010. Epigenetic differences in cytogenetically normal versus abnormal acute myeloid leukemia. *Epigenetics.* 5(7): 590-600.

Grimwade D i sar.; The Medical Research Council Adult and Children's Leukaemia Working Parties. 1998. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. *Blood.* 92(7): 2322-2333.

Grossmann V i sar. 2013. The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. *Genes Chromosomes Cancer.* 52(4): 410-422.

Grossmann V, Kohlmann A, Eder C, Haferlach C, Kern W, Cross NC, Haferlach T, Schnittger S. 2011. Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in >80% of patients with TET2 and EZH2 being of high prognostic relevance. *Leukemia.* 25(5): 877-879.

Gu Y i sar. 2003. Hematopoietic cell regulation by Rac1 and Rac2 guanosine triphosphatases. *Science.* 302(5644): 445-449.

Guo JU, Su Y, Zhong C, Ming GL, Song H. 2011. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell.* 145(3): 423-434.

Guo YZ, Li M, Lu M, Wen Z, Wang K, Li G, Wu J. 2006. Classifying G protein-coupled receptors and nuclear receptors on the basis of protein power spectrum from fast Fourier transform. *Amino Acids.* 30(4): 397-402.

Hagiwara M, Nojima T. 2007. Cross-talks between transcription and post-transcriptional events within a 'mRNA factory'. *J Biochem.* 142(1): 11-15.

Hájková H, Marková J, Haškovec C, Sárová I, Fuchs O, Kostečka A, Cetkovský P, Michalová K, Schwarz J. 2012. Decreased DNA methylation in acute myeloid leukemia patients with DNMT3A mutations and prognostic implications of DNA methylation. *Leuk Res.* 36(9): 1128-1133.

Hamburg MA, Collins FS. 2010. The path to personalized medicine. *N Engl J Med.* 363(4): 301-304.

Han Y, San-Marina S, Liu J, Minden MD. 2004. Transcriptional activation of c-myc proto-oncogene by WT1 protein. *Oncogene*. 23(41): 6933-6941.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell*. 100(1): 57-70.

Harnois T, Constantin B, Rioux A, Grenioux E, Kitzis A, Bourmeyster N. 2003. Differential interaction and activation of Rho family GTPases by p210bcr-abl and p190bcr-abl. *Oncogene*. 22(41): 6445-6454.

Hartkamp J, Carpenter B, Roberts SG. 2010. The Wilms' tumor suppressor protein WT1 is processed by the serine protease HtrA2/Omi. *Mol Cell*. 37(2): 159-171.

Hennessy BT, Smith DL, Ram PT, Lu Y, Mills GB. 2005. Exploiting the PI3K/AKT pathway for cancer drug discovery. *Nat Rev Drug Discov*. 4(12): 988-1004.

Hemsath L, Dvorsky R, Fiegen D, Carlier MF, Ahmadian MR. 2005. An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. *Mol Cell*. 20(2): 313-324.

Hermjakob H i sar. 2004. IntAct: an open source molecular interaction database. *Nucleic Acids Res*. 32(Database issue): D452-D455.

Hicks S, Wheeler DA, Plon SE, Kimmel M. 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat*. 32(6): 661-668.

Hinterleitner C, Huelsenbeck J, Henninger C, Wartlick F, Schorr A, Kaina B, Fritz G. 2013. Rac1 signaling protects monocytic AML cells expressing the MLL-AF9 oncogene from caspase-mediated apoptotic death. *Apoptosis*. 18(8): 963-979.

Ho PA i sar. 2010. Prevalence and prognostic implications of WT1 mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. *Blood*. 116(5): 702-710.

Hoelder S, Clarke PA, Workman P. 2012. Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Mol Oncol*. 6(2): 155-176.

Hoffman TL, LaBranche CC, Zhang W, Canziani G, Robinson J, Chaiken I, Hoxie JA, Doms RW. 1999. Stable exposure of the coreceptor-binding site in a CD4-independent HIV-1 envelope protein. *Proc Natl Acad Sci USA*. 96(11): 6359-6364.

Hohenstein P, Hastie ND. 2006. The many facets of the Wilms' tumour gene, WT1. *Hum Mol Genet*. 15 Spec No 2: R196-R201.

Holland SJ, Gale NW, Mbamalu G, Yancopoulos GD, Henkemeyer M, Pawson T. 1996. Bidirectional signalling through the EPH-family receptor Nuk and its transmembrane ligands. *Nature*. 383(6602): 722-725.

Holland SJ, Peles E, Pawson T, Schlessinger J. 1998. Cell-contact-dependent signalling in axon growth and guidance: Eph receptor tyrosine kinases and receptor protein tyrosine phosphatase beta. *Curr Opin Neurobiol.* 8(1): 117-127.

Hollink IH i sar. 2009. Clinical relevance of Wilms tumor 1 gene mutations in childhood acute myeloid leukemia. *Blood.* 113(23): 5951-5960.

Hoque MO i sar. 2008. Genome-wide promoter analysis uncovers portions of the cancer methylome. *Cancer Res.* 68(8): 2661-2670.

Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. 2003. Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4(10): R70.

Hosen N i sar. 2002. Very low frequencies of human normal CD34+ haematopoietic progenitor cells express the Wilms' tumour gene WT1 at levels similar to those in leukaemia cells. *Br J Haematol.* 116(2): 409-420.

Hosen N i sar. 2007. The Wilms' tumor gene WT1-GFP knock-in mouse reveals the dynamic regulation of WT1 expression in normal and leukemic hematopoiesis. *Leukemia.* 21(8): 1783-1791.

Hou HA i sar. 2010. WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system. *Blood.* 115(25): 5222-5231.

Hou HA i sar. 2012. DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. *Blood.* 119(2): 559-568.

Huang da W, Sherman BT, Lempicki RA. 2009 (a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37(1): 1-13.

Huang da W, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, Lempicki RA. 2009 (b). Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics.* Chapter 13: Unit 13.11.

Huang J, Ling CX. 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans Knowl Data Eng.* 17(3): 299-310.

Huang J, Tan PH, Li KB, Matsumoto K, Tsujimoto M, Bay BH. 2005. Y-box binding protein, YB-1, as a marker of tumor aggressiveness and response to adjuvant chemotherapy in breast cancer. *Int J Oncol.* 26(3): 607-613.

Huff V. 2011. Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. *Nat Rev Cancer.* 11(2): 111-121.

Human Genome Variation Society [Internet]. 2013. Dostupno na: <http://www.hgvs.org>

Ifeachor EC, Jarvis BW. 2002. Digital signal processing: a practical approach. Pearson Education.

Illingworth CJ, Parkes KE, Snell CR, Mullineaux PM, Reynolds CA. 2008. Criteria for confirming sequence periodicity identified by Fourier transform analysis: application to GCR2, a candidate plant GPCR? *Biophys Chem.* 133(1-3): 28-35.

Inoue K i sar. 1994. WT1 as a new prognostic factor and a new marker for the detection of minimal residual disease in acute leukemia. *Blood.* 84(9): 3071-3079.

Inoue K i sar. 1996. Long-term follow-up of minimal residual disease in leukemia patients by monitoring WT1 (Wilms tumor gene) expression levels. *Blood.* 88(6): 2267-2278.

Inoue K i sar. 1997. Aberrant overexpression of the Wilms tumor gene (WT1) in human leukemia. *Blood.* 89(4): 1405-1412.

Ionita-Laza I, Makarov V, Yoon S, Raby B, Buxbaum J, Nicolae DL, Lin X. 2011. Finding disease variants in Mendelian disorders by using sequence data: methods and applications. *Am J Hum Genet.* 89(6): 701-712.

Issa JP i sar. 2004. Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies. *Blood.* 103(5): 1635-1640.

Ito K i sar. 2006. Antiapoptotic function of 17AA(+)WT1 (Wilms' tumor gene) isoforms on the intrinsic apoptosis pathway. *Oncogene.* 25(30): 4217-4229.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science.* 333(6047): 1300-1303.

Itzykson R i sar.; Groupe Francophone des Myelodysplasies (GFM). 2011. Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias. *Leukemia.* 25(7): 1147-1152.

Ivanov AA, Khuri FR, Fu H. 2013. Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol Sci.* 34(7): 393-400.

Ivanov AS, Zgoda VG, Archakov AI. 2011. Technologies of protein interactomics: a review. *Russ J Bioorganic Chem.* 37(1): 4-16.

Ivanov TR, Jones JC, Dontenwill M, Bousquet P, Piletz JE. 1998. Characterization of a partial cDNA clone detected by imidazoline receptor-selective antisera. *J Auton Nerv Syst.* 72(2-3): 98-110.

Ivanović V, Demajo M, Todorović-Raković N, Nikolić-Vukosavljević D, Nesković-Konstantinović Z, Krtolica K, Veljković V, Prlić J, Dimitrijević B. 2004. Localization of

recognition site between transforming growth factor-beta1 (TGF-beta1) and TGF beta receptor type II: possible implications in breast cancer. *Med Hypotheses*. 62(5): 727-732.

Jain P, Baranwal S, Dong S, Struckhoff AP, Worthylake RA, Alahari SK. 2013. Integrin-binding protein nischarin interacts with tumor suppressor liver kinase B1 (LKB1) to regulate cell migration of breast epithelial cells. *J Biol Chem*. 288(22): 15495-15509.

Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR, Majeti R. 2012. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med*. 4(149) :149ra118.

Jankowska AM i sar. 2011. Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood*. 118(14): 3932-3941.

Jensen LJ i sar. 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 37(Database issue): D412-D416.

Jensen LJ, Lagarde J, von Mering C, Bork P. 2004. ArrayProspector: a web resource of functional associations inferred from microarray expression data. *Nucleic Acids Res*. 2(Web Server issue): W445-W448.

Jin L, Wessely O, Marcusson EG, Ivan C, Calin GA, Alahari SK. 2013. Prooncogenic factors miR-23b and miR-27b are regulated by Her2/Neu, EGF, and TNF- α in breast cancer. *Cancer Res*. 73(9): 2884-2896.

Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, Wang C, Brindle PK, Dent SY, Ge K. 2011. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J*. 30(2): 249-262.

Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*. 11: 431.

Johnstone RW i sar. 1996. A novel repressor, par-4, modulates transcription and growth suppression functions of the Wilms' tumor suppressor WT1. *Mol Cell Biol*. 16(12): 6945-6956.

Jonsson PF, Bates PA. 2006. Global topological features of cancer proteins in the human interactome. *Bioinformatics*. 22(18): 2291-2297.

Jordan DM, Ramensky VE, Sunyaev SR. 2010. Human allelic variation: perspective from protein function, structure, and evolution. *Curr Opin Struct Biol*. 20(3): 342-350.

Jorgensen WL. 2004. The many roles of computation in drug discovery. *Science*. 303(5665): 1813-1818.

Joshi-Tope G i sar. 2005. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res*. 33(Database issue): D428-D432.

Kaminker JS i sar. 2007. Distinguishing cancer-associated missense mutations from common polymorphisms. *Cancer Res.* 67(2): 465-473.

Kanazawa S, Okamoto T, Peterlin BM. 2000. Tat competes with CIITA for the binding to P-TEFb and blocks the expression of MHC class II genes in HIV infection. *Immunity.* 12(1): 61-70.

Kanazawa S, Soucek L, Evan G, Okamoto T, Peterlin BM. 2003. c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene.* 22(36): 5707-5711.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1): 27-30.

Kang CH, Brautigan DL, Osheroff N, Margoliash E. 1978. Definitaion of cytochrome c binding domains by chemical modification. Reaction of carboxydinitrophenyl- and trinitrophenyl-cytochromes c with baker's yeast cytochrome c peroxidase. *J Biol Chem.* 253(18): 6502-6510.

Kantarjian HM i sar. 2007. Nilotinib (formerly AMN107), a highly selective BCR-ABL tyrosine kinase inhibitor, is effective in patients with Philadelphia chromosome-positive chronic myelogenous leukemia in chronic phase following imatinib resistance and intolerance. *Blood.* 110(10): 3540-3546.

Karlsson R, Pedersen ED, Wang Z, Brakebusch C. 2009. Rho GTPase function in tumorigenesis. *Biochim Biophys Acta.* 1796(2): 91-98.

Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. 2008. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.* 36(Database issue): D202-D205.

Keilholz U, Letsch A, Busse A, Asemissen AM, Bauer S, Blau IW, Hofmann WK, Uharek L, Thiel E, Scheibenbogen C. 2009. A clinical and immunologic phase 2 trial of Wilms tumor gene product 1 (WT1) peptide vaccination in patients with AML and MDS. *Blood.* 113(26): 6541-6548.

Kelly LM, Gilliland DG. 2002. Genetics of myeloid leukemias. *Annu Rev Genomics Hum Genet.* 3: 179-198.

Kennedy D, Ramsdale T, Mattick J, Little M. 1996. An RNA recognition motif in Wilms' tumour protein (WT1) revealed by structural modelling. *Nat Genet.* 12(3): 329-331.

Kent J, Coriat AM, Sharpe PT, Hastie ND, van Heyningen V. 1995. The evolution of WT1 sequence and expression pattern in the vertebrates. *Oncogene.* 11(9): 1781-1792.

Kijima N, Hosen N, Kagawa N, Hashimoto N, Kinoshita M, Oji Y, Sugiyama H, Yoshimine T. 2014. Wilms' tumor 1 is involved in tumorigenicity of glioblastoma by regulating cell proliferation and apoptosis. *Anticancer Res.* 34(1): 61-67.

Kim H, Kang K, Kim J. 2009. AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* 37(9): 2940-2950.

Kim RH i sar. 2000. A novel smad nuclear interacting protein, SNIP1, suppresses p300-dependent TGF-beta signal transduction. *Genes Dev.* 14(13): 1605-1616.

Kim RH, Flanders KC, Birkey Reffey S, Anderson LA, Duckett CS, Perkins ND, Roberts AB. 2001. SNIP1 inhibits NF-kappa B signaling by competing for its binding to the C/H1 domain of CBP/p300 transcriptional co-activators. *J Biol Chem.* 276(49): 46297-46304.

King AD, Przulj N, Jurisica I. 2004. Protein complex prediction via cost-based clustering. *Bioinformatics.* 20(17): 3013-3020.

King-Underwood L, Pritchard-Jones K. 1998. Wilms' tumor (WT1) gene mutations occur mainly in acute myeloid leukemia and may confer drug resistance. *Blood.* 91(8): 2961-2968.

King-Underwood L, Renshaw J, Pritchard-Jones K. 1996. Mutations in the Wilms' tumor gene WT1 in leukemias. *Blood.* 87(6): 2171-2179.

Klose RJ, Bird AP. 2006. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci.* 31(2): 89-97.

Ko M i sar. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature.* 468(7325): 839-843.

Kobayashi A, Ohta T, Yamamoto M. 2004. Unique function of the Nrf2-Keap1 pathway in the inducible expression of antioxidant and detoxifying enzymes. *Methods Enzymol.* 378: 273-286.

Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet.* 39: 309-338.

Koren R, Hammes GG. 1976. A kinetic study of protein-protein interactions. *Biochemistry.* 15(5): 1165-1171.

Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. 2004. Multiple links between transcription and splicing. *RNA.* 10(10): 1489-1498.

Krsmanovic V, Biquard JM, Sikorska-Walker M, Cosic I, Desgranges C, Trabaud MA, Whitfield JF, Durkin JP, Achour A, Hearn MT. 1998. Investigations into the cross-reactivity of rabbit antibodies raised against nonhomologous pairs of synthetic peptides derived from HIV-1 gp120 proteins. *J Pept Res.* 52(5): 410-420.

Krystof V, Baumli S, Fürst R. 2012. Perspective of cyclin-dependent kinase 9 (CDK9) as a drug target. *Curr Pharm Des.* 18(20): 2883-2890.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 4(7): 1073-1081.

Kumar S, Dudley JT, Filipinski A, Liu L. 2011. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet.* 27(9): 377-386.

Ladomery M. 1997. Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. *Bioessays.* 19(10): 903-909.

Laity JH, Chung J, Dyson HJ, Wright PE. 2000. Alternative splicing of Wilms' tumor suppressor protein modulates DNA binding activity through isoform-specific DNA-induced conformational changes. *Biochemistry.* 39(18): 5341-5348.

Landwehr R, Bogdanova NV, Antonenkova N, Meyer A, Bremer M, Park-Simon TW, Hillemanns P, Karstens JH, Schindler D, Dörk T. 2011. Mutation analysis of the SLX4/FANCP gene in hereditary breast cancer. *Breast Cancer Res Treat.* 130(3): 1021-1028.

Langemeijer SM i sar. 2009. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet.* 41(7): 838-842.

LaRochelle O i sar. 2011. Do AML patients with DNMT3A exon 23 mutations benefit from idarubicin as compared to daunorubicin? A single center experience. *Oncotarget.* 2(11): 850-861.

Larsson SH, Charlier JP, Miyagawa K, Engelkamp D, Rassoulzadegan M, Ross A, Cuzin F, van Heyningen V, Hastie ND. 1995. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell.* 81(3): 391-401.

Lawrence HR i sar. 2009. Identification of a disruptor of the MDM2-p53 protein-protein interaction facilitated by high-throughput in silico docking. *Bioorg Med Chem Lett.* 19(14): 3756-3759.

Le Calvez-Kelm F i sar. 2011. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res.* 13(1): R6.

Lee TC, Lee AS, Li KB. 2008. Incorporating the amino acid properties to predict the significance of missense mutations. *Amino Acids.* 35(3): 615-626.

Lee TH, Lwu S, Kim J, Pelletier J. 2002. Inhibition of Wilms tumor 1 transactivation by bone marrow zinc finger 2, a novel transcriptional repressor. *J Biol Chem.* 277(47): 44826-44837.

Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40(Database issue): D302-D305.

Leung HC, Xiang Q, Yiu SM, Chin FY. 2009. Predicting protein complexes from PPI data: a core-attachment approach. *J Comput Biol.* 16(2): 133-144.

Leung S, Qureshi SA, Kerr IM, Darnell JE Jr, Stark GR. 1995. Role of STAT2 in the alpha interferon signaling pathway. *Mol Cell Biol.* 15(3): 1312-1317.

Ley TJ i sar. 2010. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med.* Dec 363(25): 2424-2433.

Li F, Wu N, Su R, Chen Y, Lu X, Liu Y, Li J. 2012. Imidazoline receptor antisera-selected/Nischarin regulates the effect of agmatine on the development of morphine dependence. *Addict Biol.* 17(2): 392-408.

Li F, Wu N, Su RB, Zheng JQ, Xu B, Lu XQ, Cong B, Li J. 2006. Involvement of phosphatidylcholine-selective phospholipase C in activation of mitogen-activated protein kinase pathways in imidazoline receptor antisera-selected protein. *J Cell Biochem.* 98(6): 1615-1628.

Li J, Wang Q, Fu Q, Zhu Y, Zhai Y, Yu Y, Zhang K, Yao K. 2013. A novel connexin 50 gene (gap junction protein, alpha 8) mutation associated with congenital nuclear and zonular pulverulent cataract. *Mol Vis.* 19: 767-774.

Li XL, Foo CS, Ng SK. 2007. Discovering protein complexes in dense reliable neighborhoods of protein interaction networks. *Comput Syst Bioinformatics Conf.* 6: 157-168.

Li Z i sar. 2008. Myocyte enhancer factor 2C as a neurogenic and antiapoptotic transcription factor in murine embryonic stem cells. *J Neurosci.* 28(26): 6557-6568.

Lim KP, Hong W. 2004. Human Nischarin/imidazoline receptor antisera-selected protein is targeted to the endosomes by a combined action of a PX domain and a coiled-coil region. *J Biol Chem.* 279(52): 54770-54782.

Lin J i sar. 2011. Recurrent DNMT3A R882 mutations in Chinese patients with acute myeloid leukemia and myelodysplastic syndrome. *PLoS One.* 6(10): e26906.

Liu J i sar. 2012. Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res.* 22(12): 2315-2327.

Liu L, Cai Y, Lu W, Feng K, Peng C, Niu B. 2009. Prediction of protein-protein interactions based on PseAA composition and hybrid feature selection. *Biochem Biophys Res Commun.* 380(2): 318-322.

Liu X, Jian X, Boerwinkle E. 2013. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat.* 34(9): E2393-E2402.

Loeb DM. 2006. WT1 influences apoptosis through transcriptional regulation of Bcl-2 family members. *Cell Cycle*. 5(12): 1249-1253.

LoPiccolo J, Blumenthal GM, Bernstein WB, Dennis PA. 2008. Targeting the PI3K/Akt/mTOR pathway: effective combinations and clinical considerations. *Drug Resist Updat*. 11(1-2): 32-50.

Lu Y, Nikolovska-Coleska Z, Fang X, Gao W, Shangary S, Qiu S, Qin D, Wang S. 2006. Discovery of a nanomolar inhibitor of the human murine double minute 2 (MDM2)-p53 interaction through an integrated, virtual database screening strategy. *J Med Chem*. 49(13): 3759-3762.

Lubbe SJ i sar. 2011. Evaluation of germline BMP4 mutation as a cause of colorectal cancer. *Hum Mutat*. 32(1): E1928-E1938.

Luscombe NM, Greenbaum D, Gerstein M. 2001. What is bioinformatics? An introduction and overview. *Yearb Med Inform*. 1: 83-99.

Lustig B, Behrens J. 2003. The Wnt signaling pathway and its role in tumor development. *J Cancer Res Clin Oncol*. 129(4): 199-221.

MacCallum DE, Melville J, Frame S, Watt K, Anderson S, Gianella-Borradori A, Lane DP, Green SR. 2005. Seliciclib (CYC202, R-Roscovitine) induces cell death in multiple myeloma cells by inhibition of RNA polymerase II-dependent transcription and down-regulation of Mcl-1. *Cancer Res*. 65(12): 5399-5407.

Mack NA, Whalley HJ, Castillo-Lluva S, Malliri A. 2011. The diverse roles of Rac signaling in tumorigenesis. *Cell Cycle*. 10(10): 1571-1581.

Magrane M, Consortium U. 2011 UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*. 2011: bar009.

Mahdavi MA, Lin YH. 2007. False positive reduction in protein-protein interaction predictions using gene ontology annotations. *BMC Bioinformatics*. 8: 262.

Maheswaran S i sar. 1998. Inhibition of cellular proliferation by the Wilms tumor suppressor WT1 requires association with the inducible chaperone Hsp70. *Genes Dev*. 12(8): 1108-1120.

Maheswaran S, Englert C, Bennett P, Heinrich G, Haber DA. 1995. The WT1 gene product stabilizes p53 and inhibits p53-mediated apoptosis. *Genes Dev*. 9(17): 2143-2156.

Maheswaran S, Park S, Bernard A, Morris JF, Rauscher FJ 3rd, Hill DE, Haber DA. 1993. Physical and functional interaction between WT1 and p53 proteins. *Proc Natl Acad Sci USA*. 90(11): 5100-5104.

Makki MS, Heinzl T, Englert C. 2008. TSA downregulates Wilms tumor gene 1 (Wt1) expression at multiple levels. *Nucleic Acids Res*. 36(12): 4067-4078.

Malumbres M, Barbacid M. 2005. Mammalian cyclin-dependent kinases. *Trends Biochem Sci.* 30(11): 630-641.

Mancini M, Veljkovic N, Corradi V, Zuffa E, Corrado P, Pagnotta E, Martinelli G, Barbieri E, Santucci MA. 2009. 14-3-3 ligand prevents nuclear import of c-ABL protein in chronic myeloid leukemia. *Traffic.* 10(6): 637-647.

Mancini M, Veljkovic N, Leo E, Aluigi M, Borsi E, Galloni C, Iacobucci I, Barbieri E, Santucci MA. 2012. Cytoplasmatic compartmentalization by Bcr-Abl promotes TET2 loss-of-function in chronic myeloid leukemia. *J Cell Biochem.* 113(8): 2765-2774.

Marchler-Bauer A i sar. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39(Database issue): D225-D229.

Mardis ER i sar. 2009. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med.* 361(11): 1058-1066.

Marková J i sar. 2012. Prognostic impact of DNMT3A mutations in patients with intermediate cytogenetic risk profile acute myeloid leukemia. *Eur J Haematol.* 88(2): 128-135.

Markus MA, Heinrich B, Raitskin O, Adams DJ, Mangs H, Goy C, Ladomery M, Sperling R, Stamm S, Morris BJ. 2006. WT1 interacts with the splicing protein RBM4 and regulates its ability to modulate alternative splicing in vivo. *Exp Cell Res.* 312(17): 3379-3388.

Marsella L, Sirocco F, Trovato A, Seno F, Tosatto SC. 2009. REPETITA: detection and discrimination of the periodicity of protein solenoid repeats by discrete Fourier transform. *Bioinformatics.* 25(12): i289-i295.

Martin H i sar. 2013. Pak and Rac GTPases promote oncogenic KIT-induced neoplasms. *J Clin Invest.* 123(10): 4449-4463.

Masica DL, Sosnay PR, Cutting GR, Karchin R. 2012. Phenotype-optimized sequence ensembles substantially improve prediction of disease-causing mutation in cystic fibrosis. *Hum Mutat.* 33(8): 1267-1274.

Maynadié M i sar.; HAEMACARE Working Group. 2013. Survival of European patients diagnosed with myeloid malignancies: a HAEMACARE study. *Haematologica.* 98(2): 230-238.

Mayo MW, Wang CY, Drouin SS, Madrid LV, Marshall AF, Reed JC, Weissman BE, Baldwin AS. 1999. WT1 modulates apoptosis by transcriptionally upregulating the bcl-2 proto-oncogene. *EMBO J.* 18(14): 3990-4003.

Mc Gee MM, Campiani G, Ramunno A, Nacci V, Lawler M, Williams DC, Zisterer DM. 2002. Activation of the c-Jun N-terminal kinase (JNK) signaling pathway is essential

during PBOX-6-induced apoptosis in chronic myelogenous leukemia (CML) cells. *J Biol Chem.* 277(21): 18383-18389.

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature.* 385(6614): 357-361.

McKay LM, Carpenter B, Roberts SG. 1999. Regulation of the Wilms' tumour suppressor protein transcriptional activation domain. *Oncogene.* 18(47): 6546-6554.

Meher JK, Raval MK, Dash GN, Mishra P. 2012. Prediction of secondary Structure of Proteins using Signal Processing Methods. *Int J Appl Sci Technol.* 1(1): 86-89.

Menke AL, Clarke AR, Leitch A, Ijpenberg A, Williamson KA, Spraggon L, Harrison DJ, Hastie ND. 2002. Genetic interactions between the Wilms' tumor 1 gene and the p53 gene. *Cancer Res.* 62(22): 6615-6620.

Menssen HD, Renkl HJ, Rodeck U, Maurer J, Notter M, Schwartz S, Reinhardt R, Thiel E. 1995. Presence of Wilms' tumor gene (wt1) transcripts and the WT1 nuclear protein in the majority of human acute leukemias. *Leukemia.* 9(6): 1060-1067.

Metzeler KH i sar. 2011 (a). TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol.* 29(10): 1373-1381.

Metzeler KH i sar. 2011 (b). ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. *Blood.* 118(26): 6920-6929.

Miano JM. 2003. Serum response factor: toggling between disparate programs of gene expression. *J Mol Cell Cardiol.* 35(6): 577-593.

Mikkelsen TS i sar. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* 448(7153): 553-560.

Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet.* 10(21): 2319-2328.

Mintseris J, Weng Z. 2005. Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc Natl Acad Sci USA.* 102(31): 10930-10935.

Miwa H, Beran M, Saunders GF. 1992. Expression of the Wilms' tumor gene (WT1) in human leukemias. *Leukemia.* 6(5): 405-409.

Momand J, Zambetti GP, Olson DC, George D, Levine AJ. 1992. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell.* 69(7): 1237-1245.

Morin RD i sar. 2010. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet.* 42(2): 181-185.

Morrison AA, Ladomery MR. 2006. Presence of WT1 in nuclear messenger RNP particles in the human acute myeloid leukemia cell lines HL60 and K562. *Cancer Lett.* 244(1): 136-141.

Morrison AA, Venables JP, Dellaire G, Ladomery MR. 2006. The Wilms tumour suppressor protein WT1 (+KTS isoform) binds alpha-actinin 1 mRNA via its zinc-finger domain. *Biochem Cell Biol.* 84(5): 789-798.

Morrison AA, Viney RL, Ladomery MR. 2008. The post-transcriptional roles of WT1, a multifunctional zinc-finger protein. *Biochim Biophys Acta.* 1785(1): 55-62.

Morrison DJ, English MA, Licht JD. 2005. WT1 induces apoptosis through transcriptional regulation of the proapoptotic Bcl-2 family member Bak. *Cancer Res.* 65(18): 8174-8182.

Mrózek K, Heinonen K, Bloomfield CD. 2001. Clinical importance of cytogenetics in acute myeloid leukaemia. *Best Pract Res Clin Haematol.* 14(1): 19-47.

Mukaddam-Daher S. 2012. An "I" on cardiac hypertrophic remodelling: imidazoline receptors and heart disease. *Can J Cardiol.* 28(5): 590-598.

Mukherjee P, Desai P, Zhou YD, Avery M. 2010. Targeting the BH3 domain mediated protein-protein interaction of Bcl-xL through virtual screening. *J Chem Inf Model.* 50(5): 906-923.

Müller LU, Schore RJ, Zheng Y, Thomas EK, Kim MO, Cancelas JA, Gu Y, Williams DA. 2008. Rac guanosine triphosphatases represent a potential target in AML. *Leukemia.* 22(9): 1803-1806.

Muñoz MJ, de la Mata M, Kornbliht AR. 2010. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem Sci.* 35(9): 497-504.

Murakami M. 2000. Resonant recognition model of neuropeptide Y family: hot spot amino acid distribution in the sequences. *J Protein Chem.* 19(7): 609-613.

Murati A, Brecqueville M, Devillier R, Mozziconacci MJ, Gelsi-Boyer V, Birnbaum D. 2012. Myeloid malignancies: mutations, models and management. *BMC Cancer.* 12: 304.

Murray KB, Gorse D, Thornton JM. 2002. Wavelet transforms for the characterization and detection of repeating motifs. *J Mol Biol.* 316(2): 341-363.

Murugan S, Shan J, Kühl SJ, Tata A, Pietilä I, Kühl M, Vainio SJ. 2012. WT1 and Sox11 regulate synergistically the promoter of the Wnt4 gene that encodes a critical signal for nephrogenesis. *Exp Cell Res.* 318(10): 1134-1145.

Nachtigal MW, Hirokawa Y, Enyeart-VanHouten DL, Flanagan JN, Hammer GD, Ingraham HA. 1998. Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. *Cell*. 93(3): 445-454.

Nayler O, Strätling W, Bourquin JP, Stagljar I, Lindemann L, Jasper H, Hartmann AM, Fackelmayer FO, Ullrich A, Stamm S. 1998. SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. *Nucleic Acids Res*. 26(15): 3542-3549.

Ng PC, Henikoff S. 2002. Accounting for human polymorphisms predicted to affect protein function. *Genome Res*. 12(3): 436-446.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 31(13): 3812-3814.

Nikoloski G i sar. 2010. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet*. 42(8): 665-667.

Nikolić K, Agbaba D. 2012. Ligandi imidazolinskih receptora. *Hem Ind*. 66(5): 619-635.

Nikolic K, Veljkovic N, Gemovic B, Srdic-Rajic T, Agbaba D. 2013. Imidazoline-1 receptor ligands as apoptotic agents: pharmacophore modeling and virtual docking study. *Comb Chem High Throughput Screen*. 16(4): 298-319.

Niksic M, Slight J, Sanford JR, Caceres JF, Hastie ND. 2004. The Wilms' tumour protein (WT1) shuttles between nucleus and cytoplasm and is present in functional polysomes. *Hum Mol Genet*. 13(4): 463-471.

Noble RW, Reichlin M, Gibson QH. 1969. The reactions of antibodies with heme protein antigens. The measurement of reaction kinetics and stoichiometry by fluorescence quenching. *J Biol Chem*. 244(9): 2403-2411.

Nogués G, Kadener S, Cramer P, de la Mata M, Fededa JP, Blaustein M, Srebrow A, Kornblihtt AR. 2003. Control of alternative pre-mRNA splicing by RNA Pol II elongation: faster is not always better. *IUBMB Life*. 55(4-5): 235-241.

Northrup SH, Erickson HP. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci USA*. 89(8): 3338-3342.

Nowak DE, Tian B, Jamaluddin M, Boldogh I, Vergara LA, Choudhary S, Brasier AR. 2008. RelA Ser276 phosphorylation is required for activation of a subset of NF-kappaB-dependent genes by recruiting cyclin-dependent kinase 9/cyclin T1 complexes. *Mol Cell Biol*. 28(11): 3623-3638.

Nyvold CG, Stentoft J, Braendstrup K, Melsvik D, Moestrup SK, Juhl-Christensen C, Hasle H, Hokland P. 2006. Wilms' tumor 1 mutation accumulated during therapy in acute myeloid leukemia: Biological and clinical implications. *Leukemia*. 20(11): 2051-2054.

O'Brien CE, Prideaux S, Chevassut T. 2014. The epigenetic landscape of acute myeloid leukemia. *Adv Hematol.* 2014: 103175.

Oji Y i sar. 2002. Overexpression of the Wilms' tumor gene WT1 in de novo lung cancers. *Int J Cancer.* 100(3): 297-303.

Oji Y i sar. 2003 (a). Overexpression of the Wilms' tumor gene WT1 in head and neck squamous cell carcinoma. *Cancer Sci.* 94(6): 523-529.

Oji Y i sar. 2003 (b). Overexpression of the Wilms' tumor gene WT1 in primary thyroid cancer. *Cancer Sci.* 94(7): 606-611.

Oji Y i sar. 2003 (c). Overexpression of the Wilms' tumor gene WT1 in colorectal adenocarcinoma. *Cancer Sci.* 94(8): 712-717.

Oji Y i sar. 2004 (a). Overexpression of the Wilms' tumor gene WT1 in pancreatic ductal adenocarcinoma. *Cancer Sci.* 95(7): 583-587.

Oji Y i sar. 2004 (b). Overexpression of the Wilms' tumor gene W T1 in primary astrocytic tumors. *Cancer Sci.* 95(10): 822-827.

Oji Y i sar. 2004 (c). Overexpression of the Wilms' tumor gene WT1 in esophageal cancer. *Anticancer Res.* 24(5B): 3103-3108.

Oka Y i sar. 2004. Induction of WT1 (Wilms' tumor gene)-specific cytotoxic T lymphocytes by WT1 peptide vaccine and the resultant cancer regression. *Proc Natl Acad Sci USA.* 101(38): 13885-13890.

Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell.* 99(3): 247-257.

Østergaard M, Olesen LH, Hasle H, Kjeldsen E, Hokland P. 2004. WT1 gene expression: an excellent tool for monitoring minimal residual disease in 70% of acute myeloid leukaemia patients - results from a single-centre study. *Br J Haematol.* 125(5): 590-600.

Ostrow KL i sar. 2010. Molecular analysis of plasma DNA for the early detection of lung cancer by quantitative methylation-specific PCR. *Clin Cancer Res.* 16(13): 3463-3472.

Pallant J. 2011. SPSS: Priručnik za preživljavanje. Beograd: Mikro knjiga.

Pan S, Caleshu CA, Dunn KE, Foti MJ, Moran MK, Soyinka O, Ashley EA. 2012. Cardiac structural and sarcomere genes associated with cardiomyopathy exhibit marked intolerance of genetic variation. *Circ Cardiovasc Genet.* 5(6): 602-610.

Parkinson DR, Johnson BE, Sledge GW. 2012. Making personalized cancer medicine a reality: challenges and opportunities in the development of biomarkers and companion diagnostics. *Clin Cancer Res.* 18(3): 619-624.

Paschka P i sar. 2008. Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J Clin Oncol.* 26(28): 4595-4602.

Paschka P i sar. 2010. IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *J Clin Oncol.* 28(22): 3636-3643.

Pastor WA i sar. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 473(7347): 394-397.

Patel JP i sar. 2012. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med.* 366(12): 1079-1089.

Patnaik MM, Itzykson R, Lasho TL, Kosmider O, Finke CM, Hanson CA, Knudson RA, Ketterling RP, Tefferi A, Solary E. 2014. ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: a two-center study of 466 patients. *Leukemia.* U štampi.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A.* 85(8): 2444-2448.

Peng J, Zhu Y, Milton JT, Price DH. 1998. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev.* 12(5): 755-762.

Peng SL, Gerth AJ, Ranger AM, Glimcher LH. 2001. NFATc1 and NFATc2 together control both T and B cell activation and differentiation. *Immunity.* 14(1): 13-20.

Pereira-Leal JB, Enright AJ, Ouzounis CA. 2004. Detection of functional modules from protein interaction networks. *Proteins.* 54(1): 49-57.

Perović, VR. 2013. Razvoj multifunkcionalne bioinformatičke platform zasnovane na potencijalu electron-jon interakcije bioloških molekula [disertacija]. Matematički fakultet, Univerzitet u Beogradu.

Perovic VR, Muller CP, Niman HL, Veljkovic N, Dietrich U, Tosic DD, Glisic S, Veljkovic V. 2013. Novel phylogenetic algorithm to monitor human tropism in Egyptian H5N1-HPAIV reveals evolution toward efficient human-to-human transmission. *PLoS One.* 8(4): e61572.

Peterlin BM, Price DH. 2006. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell.* 23(3): 297-305.

Peterson TA, Nehrt NL, Park D, Kann MG. 2012. Incorporating molecular and functional context into the analysis and prioritization of human variants associated with cancer. *J Am Med Inform Assoc.* 19(2): 275-283.

Phillips JE, Corces VG. 2009. CTCF: master weaver of the genome. *Cell.* 137(7): 1194-1211.

Piletz JE i sar. 2000. Imidazoline receptor antisera-selected (IRAS) cDNA: cloning and characterization. *DNA Cell Biol.* 19(6): 319-329.

Piletz JE, Jones JC, Zhu H, Bishara O, Ernsberger P. 1999. Imidazoline receptor antisera-selected cDNA clone and mRNA distribution. *Ann NY Acad Sci.* 881: 1-7.

Piletz JE, Wang G, Zhu H. 2003. Cell signaling by imidazoline-1 receptor candidate, IRAS, and the nischarin homologue. *Ann NY Acad Sci.* 1009: 392-399.

Podder S, Ghosh TC. 2010. Exploring the differences in evolutionary rates between monogenic and polygenic disease genes in human. *Mol Biol Evol.* 27(4): 934-941.

Polier G i sar. 2011. Wogonin and related natural flavones are inhibitors of CDK9 that induce apoptosis in cancer cells by transcriptional suppression of Mcl-1. *Cell Death Dis.* 2: e182.

Popowicz GM, Dömling A, Holak TA. 2011. The structure-based design of Mdm2/Mdmx-p53 inhibitors gets serious. *Angew Chem Int Ed Engl.* 50(12): 2680-2688.

Pratcorona M, Abbas S, Sanders MA, Koenders JE, Kavelaars FG, Erpelinck-Verschueren CA, Zeilemakers A, Löwenberg B, Valk PJ. 2012. Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value. *Haematologica.* 97(3): 388-392.

Price DH. 2000. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol Cell Biol.* 20(8): 2629-2634.

Prieto C, Risueño A, Fontanillo C, De las Rivas J. 2008. Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. *PLoS One.* 3(12): e3911.

Prigge JR, Iverson SV, Siders AM, Schmidt EE. 2009. Interactome for auxiliary splicing factor U2AF(65) suggests diverse roles. *Biochim Biophys Acta.* 1789(6-8): 487-492.

R Core Team. 2012. R: A language and environment for statistical computing.

Radić Z, Kirchoff PD, Quinn DM, McCammon JA, Taylor P. 1997. Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase. Distinctions between active center ligands and fasciculins. *J Biol Chem.* 272(37): 23265-23277.

Rae FK, Martinez G, Gillinder KR, Smith A, Shooter G, Forrest AR, Grimmond SM, Little MH. 2004. Analysis of complementary expression profiles following WT1 induction versus repression reveals the cholesterol/fatty acid synthetic pathways as a possible major target of WT1. *Oncogene*. 23(17): 3067-3079.

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA. 2010. c-Myc regulates transcriptional pause release. *Cell*. 141(3): 432-445.

Raje N i sar. 2005. Seliciclib (CYC202 or R-roscovitine), a small-molecule cyclin-dependent kinase inhibitor, mediates activity via down-regulation of Mcl-1 in multiple myeloma. *Blood*. 106(3): 1042-1047.

Rajendran RR, Nye AC, Frasor J, Balsara RD, Martini PG, Katzenellenbogen BS. 2003. Regulation of nuclear receptor transcriptional activity by a novel DEAD box RNA helicase (DP97). *J Biol Chem*. 278(7): 4628-4638.

Ramdass B, Chowdhary A, Koka PS. 2013. Hematological malignancies: disease pathophysiology of leukemic stem cells. *J Stem Cells*. 8(3-4): 151-187.

Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 30(17): 3894-3900.

Randjelovic J, Eric S, Savic V. 2014. In silico design of small molecule inhibitors of CDK9/cyclin T1 interaction. *J Mol Graph Model*. 50: 100-112.

Rather MI, Swamy S, Gopinath KS, Kumar A. 2014. Transcriptional repression of tumor suppressor CDC73, encoding an RNA polymerase II interactor, by Wilms tumor 1 protein (WT1) promotes cell proliferation: implication for cancer therapeutics. *J Biol Chem*. 289(2): 968-976.

Reddig PJ, Xu D, Juliano RL. 2005. Regulation of p21-activated kinase-independent Rac1 signal transduction by nischarin. *J Biol Chem*. 280(35): 30994-31002.

Refaat MM i sar. 2012. Genetic variation in the alternative splicing regulator RBM20 is associated with dilated cardiomyopathy. *Heart Rhythm*. 9(3): 390-396.

Reizner N, Maor S, Sarfstein R, Abramovitch S, Welshons WV, Curran EM, Lee AV, Werner H. 2005. The WT1 Wilms' tumor suppressor gene product interacts with estrogen receptor-alpha and regulates IGF-I receptor gene transcription in breast cancer cells. *J Mol Endocrinol*. 35(1): 135-144.

Renneville A i sar. 2012. Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia: a study by the Acute Leukemia French Association. *Leukemia*. 26(6): 1247-1254.

Renneville A, Boissel N, Zurawski V, Llopis L, Biggio V, Nibourel O, Philippe N, Thomas X, Dombret H, Preudhomme C. 2009. Wilms tumor 1 gene mutations are

associated with a higher risk of recurrence in young adults with acute myeloid leukemia: a study from the Acute Leukemia French Association. *Cancer*. 115(16): 3719-3727.

Renneville A, Roumier C, Biggio V, Nibourel O, Boissel N, Fenaux P, Preudhomme C. 2008. Cooperating gene mutations in acute myeloid leukemia: a review of the literature. *Leukemia*. 22(5): 915-931.

Rew Y i sar. 2012. Structure-based design of novel inhibitors of the MDM2-p53 interaction. *J Med Chem*. 55(11): 4936-4954.

Rezvani K, Yong AS, Mielke S, Savani BN, Musse L, Superata J, Jafarpour B, Boss C, Barrett AJ. 2008. Leukemia-associated antigen-specific T-cell responses following combined PR1 and WT1 peptide vaccination in patients with myeloid malignancies. *Blood*. 111(1): 236-242.

Rhee I i sar. 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*. 416(6880): 552-556.

Richard DJ, Schumacher V, Royer-Pokora B, Roberts SG. 2001. Par4 is a coactivator for a splice isoform-specific transcriptional activation domain in WT1. *Genes Dev*. 15(3): 328-339.

Ristola M, Arpiainen S, Saleem MA, Holthöfer H, Lehtonen S. 2012. Transcription of nephrin-Neph3 gene pair is synergistically activated by WT1 and NF- κ B and silenced by DNA methylation. *Nephrol Dial Transplant*. 27(5): 1737-1745.

Risueño A, Fontanillo C, Dinger ME, De Las Rivas J. 2010 GATEExplorer: genomic and transcriptomic explorer; mapping expression probes to gene loci, transcripts, exons and ncRNAs. *BMC Bioinformatics*. 11: 221.

Roberts SG. 2006. The modulation of WTI transcription function by cofactors. *Biochem Soc Symp*. (73): 191-201.

Robinson V, Smith A, Flenniken AM, Wilkinson DG. 1997. Roles of Eph receptors and ephrins in neural crest pathfinding. *Cell Tissue Res*. 290(2): 265-274.

Roslan R, Othman RM, Shah ZA, Kasim S, Asmuni H, Taliba J, Rohayanti H, Zakaria Z. 2010. Incorporating multiple genomic features with the utilization of interacting domain patterns to improve the prediction of protein–protein interactions. *Inf Sci*. 180(20): 3955-3973.

Rossi D i sar. 2011. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*. 118(26): 6904-6908.

Rountree MR, Bachman KE, Baylin SB. 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet*. 25(3): 269-277.

Ryan DP, Matthews JM. 2005. Protein-protein interactions in human disease. *Curr Opin Struct Biol.* 15(4): 441-446.

Sano H, Liu SC, Lane WS, Piletz JE, Lienhard GE. 2002. Insulin receptor substrate 4 associates with the protein IRAS. *J Biol Chem.* 277(22): 19439-19447.

Sant M i sar.; HAEMACARE Working Group. 2010. Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. *Blood.* 116(19): 3724-3734.

Santo L i sar. 2010. AT7519, A novel small molecule multi-cyclin-dependent kinase inhibitor, induces apoptosis in multiple myeloma via GSK-3beta activation and RNA polymerase II inhibition. *Oncogene.* 29(16): 2325-2336.

Satijn DP, Gunster MJ, van der Vlag J, Hamer KM, Schul W, Alkema MJ, Saurin AJ, Freemont PS, van Driel R, Otte AP. 1997. RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. *Mol Cell Biol.* 17(7): 4105-4113.

Sauren M, Pirogova E, Cosic I. 2004. RRM analysis of protoporphyrinogen oxidase. *Australas Phys Eng Sci Med.* 27(4): 174-179.

Scharnhorst V, Dekker P, van der Eb AJ, Jochemsen AG. 2000. Physical interaction between Wilms tumor 1 and p73 proteins modulates their functions. *J Biol Chem.* 275(14): 10202-10211.

Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Müller J. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature.* 465(7295): 243-247.

Schlosshauer M, Baker D. 2004. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Sci.* 13(6): 1660-1669.

Schnittger S i sar. 2013. ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. *Leukemia.* 27(1): 82-91.

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell.* 128(4): 735-745.

Scott JW, Hawley SA, Green KA, Anis M, Stewart G, Scullion GA, Norman DG, Hardie DG. 2004. CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J Clin Invest.* 113(2): 274-284.

Sekeres MA i sar. 2012. Phase 2 study of the lenalidomide and azacitidine combination in patients with higher-risk myelodysplastic syndromes. *Blood.* 120(25): 4945-4951.

Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, Sawyers CL. 2002. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell*. 2(2): 117-125.

Shen HB, Chou KC. 2006. Ensemble classifier for protein fold pattern recognition. *Bioinformatics*. 22(14): 1717-1722.

Shen HB, Chou KC. 2008. PseAAC: a flexible web server for generating various kinds of protein pseudo amino acid composition. *Anal Biochem*. 373(2): 386-388.

Shen J, Zhang J, Luo X, Zhu W, Yu K, Chen K, Li Y, Jiang H. 2007. Predicting protein-protein interactions based only on sequences information. *Proc Natl Acad Sci USA*. 104(11): 4337-4341.

Shen Y i sar. 2011. Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood*. 118(20): 5593-5603.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 29(1): 308-311.

Shihab HA, Gough J, Cooper DN, Day IN, Gaunt TR. 2013 (a). Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics*. 29(12): 1504-1510.

Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, Gaunt TR. 2013 (b). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 34(1): 57-65.

Shih AH, Abdel-Wahab O, Patel JP, Levine RL. 2012. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer*. 12(9): 599-612.

Shoemaker BA, Panchenko AR. 2007. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol*. 3(4): e43.

Sebro R, Levy H, Schneck K, Dimmock D, Raby BA, Cannon CL, Broeckel U, Risch NJ. 2012. Cystic fibrosis mutations for p.F508del compound heterozygotes predict sweat chloride levels and pancreatic sufficiency. *Clin Genet*. 82(6): 546-551.

Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Krämer A, Sattler M. 2003. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell*. 11(4): 965-976.

Shain AH, Pollack JR. 2013. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One*. 8(1): e55119.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*. 479(7371): 74-79.

Silverman LR i sar. 2002. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol*. 20(10): 2429-2440.

Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 40(Web Server issue): W452-W457.

Simon JA, Kingston RE. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol*. 10(10): 697-708.

Simpson LA, Burwell EA, Thompson KA, Shahnaz S, Chen AR, Loeb DM. 2006. The antiapoptotic gene A1/BFL1 is a WT1 target gene that mediates granulocytic differentiation and resistance to chemotherapy. *Blood*. 107(12): 4695-4702.

Skerl V, Pavlović M. 1988. Thymopoietins and long postsynaptic neurotoxins share common information in their primary structures. *FEBS Lett*. 239(1): 141-146.

Skorski T, Wlodarski P, Daheron L, Salomoni P, Nieborowska-Skorska M, Majewski M, Wasik M, Calabretta B. 1998. BCR/ABL-mediated leukemogenesis requires the activity of the small GTP-binding protein Rac. *Proc Natl Acad Sci USA*. 95(20): 11858-11862.

Slovak ML i sar. 2000. Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood*. 96(13): 4075-4083.

Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, Copeland RA. 2010. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A*. 107(49): 20980-20985.

Somervaille TC, Cleary ML. 2006. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell*. 10(4): 257-268.

Spirin V, Mirny LA. 2003. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA*. 100(21): 12123-12128.

Srichai MB i sar. 2004. A WT1 co-regulator controls podocyte phenotype by shuttling between adhesion structures and nucleus. *J Biol Chem*. 279(14): 14398-14408.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*. 34(Database issue): D535-D539.

Stegelmann F, Bullinger L, Schlenk RF, Paschka P, Griesshammer M, Blersch C, Kuhn S, Schauer S, Döhner H, Döhner K. 2011. DNMT3A mutations in myeloproliferative neoplasms. *Leukemia*. 25(7): 1217-1219.

Stehling-Sun S, Dade J, Nutt SL, DeKoter RP, Camargo FD. 2009. Regulation of lymphoid versus myeloid fate 'choice' by the transcription factor Mef2c. *Nat Immunol*. 10(3): 289-296.

Stephens P i sar. 2005. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet*. 37(6): 590-592.

Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher AG, Pombo A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol*. 9(12): 1428-1435.

Stojković L, Djurić T, Stanković A, Dinčić E, Stančić O, Veljković N, Alavantić D, Zivković M. 2012. The association of V249I and T280M fractalkine receptor haplotypes with disease course of multiple sclerosis. *J Neuroimmunol*. 245(1-2): 87-92.

Stone EA, Sidow A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res*. 15(7): 978-986.

Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature*. 458(7239): 719-724.

Summers K i sar. 2007. Wilms' tumour 1 mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia*. 21(3): 550-551.

Sun Z, Chang CH, Ernsberger P. 2007. Identification of IRAS/Nischarin as an II-imidazoline receptor in PC12 rat pheochromocytoma cells. *J Neurochem*. 101(1): 99-108.

Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN. 1999. PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng*. 12(5): 387-394.

Sutherland KD, Lindeman GJ, Choong DY, Wittlin S, Brentzell L, Phillips W, Campbell IG, Visvader JE. 2004. Differential hypermethylation of SOCS genes in ovarian and breast carcinomas. *Oncogene*. 23(46): 7726-7733.

Szklarczyk D i sar. 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 39(Database issue): D561-D568.

Tahiliani M i sar. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 324(5929): 930-935.

Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. 2010. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature*. 465(7299): 747-751.

Talpaz M i sar. 2006. Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias. *N Engl J Med*. 354(24): 2531-2541.

Tantin D, Schild-Poulter C, Wang V, Haché RJ, Sharp PA. 2005. The octamer binding transcription factor Oct-1 is a stress sensor. *Cancer Res*. 65(23): 10750-10758.

Tatsumi N i sar. 2008. Wilms' tumor gene WT1-shRNA as a potent apoptosis-inducing agent for solid tumors. *Int J Oncol*. 32(3): 701-711.

Tchernitchko D, Goossens M, Wajcman H. 2004. In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. *Clin Chem*. 50(11): 1974-1978.

Tefferi A i sar. 2010. IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis. *Leukemia*. 24(7): 1302-1309.

Tefferi A, Vainchenker W. 2011. Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. *J Clin Oncol*. 29(5): 573-582.

Tesfai J, Crane L, Baziard-Mouysset G, Kennedy W, Edwards LP. 2011 Novel II-imidazoline S43126 enhance insulin action in PC12 cells. *Pharmacol Rep*. 63(6): 1442-1449.

Thiel G, Cibelli G. 2002. Regulation of life and death by the zinc finger transcription factor Egr-1. *J Cell Physiol*. 193(3): 287-292.

Thol F i sar. 2011 (a). Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J Clin Oncol*. 29(21): 2889-2896.

Thol F i sar. 2011 (b). Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes. *J Clin Oncol*. 29(18): 2499-2506.

Thol F, Winschel C, Lüdeking A, Yun H, Friesen I, Damm F, Wagner K, Krauter J, Heuser M, Ganser A. 2011 (c). Rare occurrence of DNMT3A mutations in myelodysplastic syndromes. *Haematologica*. 96(12): 1870-1873.

Thomas EK i sar. 2007. Rac guanosine triphosphatases represent integrating molecular therapeutic targets for BCR-ABL-induced myeloproliferative disease. *Cancer Cell*. 12(5): 467-478.

Thomas EK, Cancelas JA, Zheng Y, Williams DA. 2008. Rac GTPases as key regulators of p210-BCR-ABL-dependent leukemogenesis. *Leukemia*. 22(5): 898-904.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13(9): 2129-2141.

Thomas PD, Kejariwal A. 2004. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci USA.* 101(43): 15398-15403.

Thompson T, Andreeff M, Studzinski GP, Vassilev LT. 2010. 1,25-dihydroxyvitamin D3 enhances the apoptotic activity of MDM2 antagonist nutlin-3a in acute myeloid leukemia cells expressing wild-type p53. *Mol Cancer Ther.* 9(5): 1158-1168.

Thusberg J, Olatubosun A, Vihinen M. 2011. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 32(4): 358-368.

Tintori C i sar. 2007. Novel virtual screening protocol based on the combined use of molecular modeling and electron-ion interaction potential techniques to design HIV-1 integrase inhibitors. *J Chem Inf Model.* 47(4): 1536-1544.

Tintori C, Veljkovic N, Veljkovic V, Botta M. 2010. Computational studies of the interaction between the HIV-1 integrase tetramer and the cofactor LEDGF/p75: insights from molecular dynamics simulations and the informational spectrum method. *Proteins.* 78(16): 3396-3408.

Torkamani A, Schork NJ. 2007. Accurate prediction of deleterious protein kinase polymorphisms. *Bioinformatics.* 23(21): 2918-2925.

Torkamani A, Verkhivker G, Schork NJ. 2009. Cancer driver mutations in protein kinase genes. *Cancer Lett.* 281(2): 117-127.

Tsuboi A i sar. 1999. Constitutive expression of the Wilms' tumor gene WT1 inhibits the differentiation of myeloid progenitor cells but promotes their proliferation in response to granulocyte-colony stimulating factor (G-CSF). *Leuk Res.* 23(5): 499-505.

Tuna M, Chavez-Reyes A, Tari AM. 2005. HER2/neu increases the expression of Wilms' Tumor 1 (WT1) protein to stimulate S-phase proliferation and inhibit apoptosis in breast cancer cells. *Oncogene.* 24(9): 1648-1652.

Ueda T i sar. 2003. Overexpression of the Wilms' tumor gene WT1 in human bone and soft-tissue sarcomas. *Cancer Sci.* 94(3): 271-276.

Uesugi K, Hiasa Y, Tokumoto Y, Mashiba T, Koizumi Y, Hirooka M, Abe M, Matsuura B, Onji M. 2013. Wilms' tumor 1 gene modulates Fas-related death signals and anti-apoptotic functions in hepatocellular carcinoma. *J Gastroenterol.* 48(9): 1069-1080.

UniProt Consortium. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42(Database issue): D191-D198.

Valinluck V, Sowers LC. 2007. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* 67(3): 946-950.

van Blitterswijk M, van Vught PW, van Es MA, Schelhaas HJ, van der Kooi AJ, de Visser M, Veldink JH, van den Berg LH. 2012. Novel optineurin mutations in sporadic amyotrophic lateral sclerosis patients. *Neurobiol Aging.* 33(5): 1016.e1-1016.e 7.

Varambally S i sar. 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature.* 419(6907): 624-629.

Varambally S i sar. 2008. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science.* 322(5908): 1695-1699.

Vardiman JW. 2010. The World Health Organization (WHO) classification of tumors of the hematopoietic and lymphoid tissues: an overview with emphasis on the myeloid neoplasms. *Chem Biol Interact.* 184(1-2): 16-20.

Vardiman JW i sar. 2009. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood.* 114(5): 937-951.

Vassilev LT i sar. 2004. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science.* 303(5659): 844-848.

Veljkovic N, Branch DR, Metlas R, Prljic J, Vlahovicek K, Pongor S, Veljkovic V. 2003. Design of peptide mimetics of HIV-1 gp120 for prevention and therapy of HIV disease. *J Pept Res.* 62(4): 158-166.

Veljkovic N, Perovic V. 2006. In silico criterion for prediction of effects of p53 gene missense mutations on p53-Mdm2 feedback loop. *Protein Pept Lett.* 13(8): 807-814.

Veljkovic, V. 1980. A theoretical approach to preselection of carcinogens and chemical carcinogenesis. New York: Gordon & Breach.

Veljković V, Cosić I, Dimitrijević B, Lalović D. 1985. Is it possible to analyze DNA and protein sequences by the methods of digital signal processing? *IEEE Trans Biomed Eng.* 32(5): 337-341.

Veljković V, Metlas R. 1988. Identification of nanopeptide from HTLV-III, ARV-2 and LAVBRU envelope gp120 determining binding to T4 cell surface protein. *Cancer Biochem Biophys.* 10(2): 91-106.

Veljkovic V, Niman HL, Glisic S, Veljkovic N, Perovic V, Muller CP. 2009 (a). Identification of hemagglutinin structural domain and polymorphisms which may modulate swine H1N1 interactions with human receptor. *BMC Struct Biol.* 9: 62.

Veljković V, Slavić I. 1972. Simple general-model pseudopotential. *Phys Rev Lett.* 29(2): 105.

Veljkovic V, Veljkovic N, Esté JA, Hüther A, Dietrich U. 2007. Application of the EIIP/ISM bioinformatics concept in development of new drugs. *Curr Med Chem.* 14(4): 441-453.

Veljkovic V, Veljkovic N, Muller CP, Müller S, Glisic S, Perovic V, Köhler H. 2009 (b). Characterization of conserved properties of hemagglutinin of H5N1 and human influenza viruses: possible consequences for therapy and infection control. *BMC Struct Biol.* 9: 21.

Vemula SR, Puschmann A, Xiao J, Zhao Y, Rudzińska M, Frei KP, Truong DD, Wszolek ZK, LeDoux MS. 2013. Role of $G\alpha(\text{olf})$ in familial and sporadic adult-onset primary dystonia. *Hum Mol Genet.* 22(12): 2510-2519.

Villa R, De Santis F, Gutierrez A, Minucci S, Pelicci PG, Di Croce L. 2004. Epigenetic gene silencing in acute promyelocytic leukemia. *Biochem Pharmacol.* 68(6): 1247-1254.

Virappane P i sar. 2008. Mutation of the Wilms' tumor 1 gene is a poor prognostic factor associated with chemotherapy resistance in normal karyotype acute myeloid leukemia: the United Kingdom Medical Research Council Adult Leukaemia Working Party. *J Clin Oncol.* 26(33): 5429-5435.

Visser O i sar.; RARECARE Working Group. 2012. Incidence, survival and prevalence of myeloid malignancies in Europe. *Eur J Cancer.* 48(17): 3257-3266.

von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. 2005. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33(Database issue): D433-D437.

Vu BT, Vassilev L. 2011. Small-molecule inhibitors of the p53-MDM2 interaction. *Curr Top Microbiol Immunol.* 348: 151-172.

Wagle N i sar. 2011. Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol.* 29(22): 3085-3096.

Walter MJ i sar. 2011. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia.* 25(7): 1153-1158.

Walter MJ i sar. 2012. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med.* 366(12): 1090-1098.

Wang CY, Xu Q, Weng L, Zhang Q, Zhang HN, Guo JF, Tan LM, Tang JG, Yan XX, Tang BS. 2011. Genetic variations of Omi/HTRA2 in Chinese patients with Parkinson's disease. *Brain Res.* 1385: 293-297.

Wang H i sar. 2010. Mutations in NEXN, a Z-disc gene, are associated with hypertrophic cardiomyopathy. *Am J Hum Genet.* 87(5): 687-693.

Wang S, Fischer PM. 2008. Cyclin-dependent kinase 9: a key transcriptional regulator and potential drug target in oncology, virology and cardiology. *Trends Pharmacol Sci.* 29(6): 302-313.

Wang W, Lee SB, Palmer R, Ellisen LW, Haber DA. 2001. A functional interaction with CBP contributes to transcriptional activation by the Wilms tumor suppressor WT1. *J Biol Chem.* 276(20): 16810-16816.

Wang X i sar. 2013. Wilms' tumour suppressor gene 1 (WT1) is involved in the carcinogenesis of Lung cancer through interaction with PI3K/Akt pathway. *Cancer Cell Int.* 13(1): 114.

Wang Y i sar. 2009. LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell.* 138(4): 660-672.

Wang ZY, Qiu QQ, Deuel TF. 1993. The Wilms' tumor gene product WT1 activates or suppresses transcription through separate functional domains. *J Biol Chem.* 268(13): 9172-9175.

Wang ZY, Qiu QQ, Seufert W, Taguchi T, Testa JR, Whitmore SA, Callen DF, Welsh D, Shenk T, Deuel TF. 1996. Molecular cloning of the cDNA and chromosome localization of the gene for human ubiquitin-conjugating enzyme 9. *J Biol Chem.* 271(40): 24811-24816.

Watanabe N, Narita M, Yamahira A, Taniguchi T, Furukawa T, Yoshida T, Miyazawa T, Nashimoto M, Takahashi M. 2013. Induction of apoptosis of leukemic cells by TRUE gene silencing using small guide RNAs targeting the WT1 mRNA. *Leuk Res.* 37(5): 580-585.

Weisberg E, Boulton C, Kelly LM, Manley P, Fabbro D, Meyer T, Gilliland DG, Griffin JD. 2002. Inhibition of mutant FLT3 receptors in leukemia cells by the small molecule tyrosine kinase inhibitor PKC412. *Cancer Cell.* 1(5): 433-443.

Weisser M, Kern W, Rauhut S, Schoch C, Hiddemann W, Haferlach T, Schnittger S. 2005. Prognostic impact of RT-PCR-based quantification of WT1 gene expression during MRD monitoring of acute myeloid leukemia. *Leukemia.* 19(8): 1416-1423.

Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA. 1998. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl Chem.* 70(5): 1129-1143.

Whisstock JC, Lesk AM. 2003. Prediction of protein function from protein sequence and structure. *Q Rev Biophys.* 36(3): 307-340.

Widschwendter M i sar. 2007. Epigenetic stem cell signature in cancer. *Nat Genet.* 39(2): 157-158.

Wilkinson DG. 2001. Multiple roles of EPH receptors and ephrins in neural development. *Nat Rev Neurosci.* 2(3): 155-164.

Wu M, Li X, Kwoh CK, Ng SK. 2009. A core-attachment based method to detect protein complexes in PPI networks. *BMC Bioinformatics.* 10: 169.

Wu N, Su RB, Liu Y, Lu XQ, Zheng JQ, Cong B, Li J. 2006. Modulation of agmatine on calcium signal in morphine-dependent CHO cells by activation of IRAS, a candidate for imidazoline I1 receptor. *Eur J Pharmacol.* 548(1-3): 21-28.

Wu X, Xu W, Cui G, Yan Y, Wu X, Li L, Tan X, Wu Q, Gu X. 2013. The expression pattern of Nischarin after lipopolysaccharides (LPS)-induced neuroinflammation in rats brain cortex. *Inflamm Res.* 62(11): 929-940.

Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. 2002. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 30(1): 303-305.

Xiong H, Buckwalter BL, Shieh HM, Hecht MH. 1995. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proc Natl Acad Sci USA.* 92(14): 6349-6353.

Xu B, Zeng DQ, Wu Y, Zheng R, Gu L, Lin X, Hua X, Jin GH. 2011. Tumor suppressor menin represses paired box gene 2 expression via Wilms tumor suppressor protein-polycomb group complex. *J Biol Chem.* 286(16): 13937-13944.

Xu C i sar. 2013. WT1 promotes cell proliferation in non-small cell lung cancer cell lines through up-regulating cyclin D1 and p-pRb in vitro and in vivo. *PLoS One.* 8(8): e68837.

Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W. 1998. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell.* 2(6): 851-861.

Yamagami T, Sugiyama H, Inoue K, Ogawa H, Tatekawa T, Hirata M, Kudoh T, Akiyama T, Murakami A, Maekawa T. 1996. Growth inhibition of human leukemic cells by WT1 (Wilms tumor gene) antisense oligodeoxynucleotides: implications for the involvement of WT1 in leukemogenesis. *Blood.* 87(7): 2878-2884.

Yan H i sar. 2009. IDH1 and IDH2 mutations in gliomas. *N Engl J Med.* 360(8): 765-773.

Yan XJ i sar. 2011. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet.* 43(4): 309-315.

Yang L, Han Y, Suarez Saiz F, Minden MD. 2007. A tumor suppressor and oncogene: the WT1 story. *Leukemia*. 21(5): 868-876.

Yang SY. 2010. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today*. 15(11-12): 444-450.

Yue P, Forrest WF, Kaminker JS, Lohr S, Zhang Z, Cavet G. 2010. Inferring the functional effects of mutation through clusters of mutations in homologous proteins. *Hum Mutat*. 31(3): 264-271.

Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G. 2002. MINT: a Molecular INTERaction database. *FEBS Lett*. 513(1): 135-140.

Zhang DD, Hannink M. 2003. Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. *Mol Cell Biol*. 23(22): 8137-8151.

Zhang G, Spellman DS, Skolnik EY, Neubert TA. 2006. Quantitative phosphotyrosine proteomics of EphB2 signaling by stable isotope labeling with amino acids in cell culture (SILAC). *J Proteome Res*. 5(3): 581-588.

Zhang J, Abdel-Rahman AA. 2008. Inhibition of nischarin expression attenuates rilmenidine-evoked hypotension and phosphorylated extracellular signal-regulated kinase 1/2 production in the rostral ventrolateral medulla of rats. *J Pharmacol Exp Ther*. 324(1): 72-78.

Zhang S, Wang T. 2009. Feature analysis of protein structure by using discrete Fourier transform and continuous wavelet transform. *J Math Chem*. 46(2): 562-568.

Zhao S i sar. 2009. Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. *Science*. 324(5924): 261-265.

Zhou HX. 1997. Enhancement of protein-protein association rate by interaction potential: accuracy of prediction based on local Boltzmann factor. *Biophys J*. 73(5): 2441-2445.

Zhuang C i sar. 2012. Discovery, synthesis, and biological evaluation of orally active pyrrolidone derivatives as novel inhibitors of p53-MDM2 protein-protein interaction. *J Med Chem*. 55(22): 9630-9642.

Zinzalla G, Thurston DE. 2009. Targeting protein-protein interactions for therapeutic intervention: a challenge for the future. *Future Med Chem*. 1(1): 65-93.

Zorio DA, Bentley DL. 2004. The link between mRNA processing and transcription: communication works both ways. *Exp Cell Res*. 296(1): 91-97.

Prilog 1

LISTA1: Lista proteinskih kandidata dobijenih kros-spektralnom analizom proteina WT1.

(Primena integrativnog algoritma za predikciju PPI na protein WT1)

P07197	O15409	Q496J9	Q96SM3	Q8TE02	Q9P0U3	Q9H3D4
Q86U42	P07199	Q9H0F6	Q6UY11	Q9UBU9	P28223	Q5T7W0
Q05519	O15496	Q92908	O00634	Q8NHJ6	P98082	Q9C0K0
Q9H307	Q9Y2V7	P31277	Q86X29	Q92934	Q13433	O75355
Q96QF7	O60828	O60906	Q8IYR2	P14672	Q13617	Q8N0W7
Q9BTD8	Q12946	Q9UJM8	Q969F2	Q9Y6J9	Q6N063	Q9H089
Q9BUV0	Q96NS5	Q7L3S4	Q8WWP7	P14207	A1L443	Q9P0L9
Q96IZ7	Q6NYC1	Q9UIK4	P0C862	Q9BRZ2	P60508	Q5T2D2
Q15428	Q9NVA4	P61019	P18669	Q9UHI7	Q8NHY3	Q6ISB3
Q8TBK6	Q8NEA6	Q96AJ1	Q9H4Q3	P52849	Q8N2C3	P62191
Q9NWB6	Q9UIH9	P35247	Q8NB78	P36957	Q9BYG4	Q9UJH3
Q9HCD5	Q9UBC3	Q9BUJ2	Q7Z3G6	Q9P0V8	P55735	P57059
O95218	Q99583	Q8NEY8	O96011	P00739	Q8TAD8	Q5VU13
Q5BKZ1	Q8IXQ8	Q6Q0C0	Q0GE19	Q9P1W8	P08908	Q86UC2
O95232	P48730	Q13620	Q7Z7M1	P11509	Q96AA8	P49862
Q16630	Q9BY84	O00571	Q9BQY4	Q96EW2	Q8NCL9	Q9NQS3
P23588	P84996	P51813	P18850	Q13087	Q2TAA8	Q8N1E2
Q15287	Q2M243	P57768	A8MU46	P27449	Q969T3	Q6UXN2
Q13523	P10073	Q8NDH8	Q05C16	Q4VXF1	Q52LR7	Q8IXT1
Q9NZ81	Q96CK0	O96017	P78352	P49590	Q96K49	Q8N0X2
Q8N7U7	Q9Y2L6	Q1X8D7	Q8N3Z6	P78383	O95859	Q9UL18
Q9Y2W2	P98175	P53539	Q9NZW5	Q8TCS8	Q8NB15	Q9UDX4
P20849	Q16537	Q96P69	P48775	Q86V42	Q0VDG4	P24387
Q8TF01	O94880	Q96EY9	Q9Y6F8	Q6P2D8	Q9BYC5	P07510
A7MD48	P56975	Q96BZ8	Q8TDW4	Q9UHF1	P14923	O43638
Q9BXN1	P25929	Q9UL49	P12830	Q68DV7	O15269	Q9UK59
Q15696	Q16384	Q96A28	O00167	Q8NGE7	Q8NHK3	Q9ULQ0
P04632	Q9Y240	Q13950	P30279	Q96MU8	P25103	A0AVK6
Q16629	P21964	Q9UNH7	Q8N568	Q9BQ67	Q9Y663	Q49AN0
P50552	P12035	Q02078	Q99679	Q9H6Q4	Q96LW9	Q96BJ3
Q12999	Q5TZF3	P0C7U0	P18577	Q8N8F6	Q8IY34	Q13478
P18615	Q6P7N7	Q70CQ3	A6NDG6	Q6UXU4	Q01892	P30508
Q9Y580	Q8N755	P05160	Q6IQ49	P26436	P20941	P38405
Q9ULX6	Q76M96	Q96GA3	O94993	Q9BSW7	Q07960	Q9NP60
O43763	Q5JXM2	P62079	Q15904	Q9H4G0	P13646	Q9BZB8
Q8IWT6	O14763	Q9NRL3	P03886	Q9UKJ0	Q86W26	Q9NRX5
Q9UBV8	P13378	Q96DN6	Q15415	P14859	Q8IZJ6	Q9UBN6
Q15695	Q8TDN4	Q9BUL5	P21854	Q9H8Y5	Q8TAA5	Q13685
A6NJI9	Q8N4U5	Q9Y6W5	Q75N03	Q96K62	Q13568	Q9NVM4
Q8NDX5	O14958	Q9BYT9	Q03014	Q99944	Q9NR34	P30048
P49918	B2RU33	P17022	O60869	O75952	Q9NZL9	P09466
P49760	Q5SQQ9	Q02446	Q8TBN0	Q86UP9	Q96PR1	Q8IVT5
Q96SZ4	Q9NZT2	Q5H9L4	Q9HCJ3	Q9Y2K5	Q8NEN0	P50402
Q8WVK2	Q9Y2K6	Q8IZU9	P56270	A7XYQ1	Q9Y5B0	Q8IUM7
Q6X4W1	O43759	Q9NQT4	Q5HYW3	Q04609	Q96SQ7	P30613
Q5VTL8	Q14498	Q6P461	Q8NH81	Q96I25	Q9BS92	Q9UL63

Q8ND83	P48668	Q93097	Q8ND24	O75071	Q9HCQ7	Q8TDG2
P31314	Q9Y692	P08912	Q96RL6	Q92832	Q96DM1	Q9NVN8
P84103	Q9UBS3	Q9BRD0	Q9NQZ6	P24723	Q6VB84	P50995
P98169	P08709	Q969T4	A6NFU0	Q9BWU1	Q5VV16	Q9BVG4
Q5SSG8	Q13427	Q9H9D4	P19419	P58872	Q9NS82	P61247
Q9BXP5	O43313	Q86Z23	Q9NY43	P56703	P29083	Q8N5W9
Q8NAV1	Q6E0U4	P20591	Q9H814	Q9H7E2	Q7RTT4	Q32MZ4
Q13516	P42331	Q96QA5	Q96GI7	Q8N165	P32302	Q8N1G4
Q95251	Q9Y2Y4	Q9Y6M4	Q86XN8	Q9Y253	O95376	Q96T91
Q7L4I2	Q15517	P15291	P10275	P46089	Q8IYG6	Q7Z5S9
Q6L8H2	O75954	O43318	Q9NY84	Q9Y3B9	Q8WUT4	Q9H8E8
Q99453	P58012	Q8WZ75	Q8WXF0	O95905	Q6NUP7	Q9Y5F6
Q66PJ3	Q9NQL9	P30626	Q6ZTI6	P47804	Q5MIZ7	Q8TE69
Q96A59	Q08AE8	Q15464	P49715	O95684	P30085	P11161
Q9UK58	O75525	Q4G1C9	O60304	Q12980	Q8N143	Q8WV83
Q13285	Q8NG27	Q8N4S9	Q96B97	Q6ZMJ2	Q6QHF9	P56746
Q6UW01	Q8NA70	Q9BY21	Q8WTX9	Q9Y5P4	P08294	Q9Y231
Q9H8W4	Q9H0K4	O00213	Q6S5H5	P14060	Q5VUG0	P30505
Q9UJY1	Q8NEF9	O00322	P22223	Q96CH1	P05186	P22309
O43462	Q9HCS4	P50750	Q8NH59	Q8N4E4	P04628	Q96DE9
P38159	P33032	P09758	Q30201	P56159	Q8N1S5	Q15572
Q9BQA5	O60237	Q7Z4L9	Q9UBL3	P07384	Q9Y3A5	P08620
O75190	Q8NEP3	P48960	Q6IN85	Q5VZ66	P30501	Q9HCP0
Q8IYL3	Q8NBI3	Q9BXJ2	Q15942	O43915	Q96FW1	Q7Z6J6
P55316	A6NI47	P13584	P17480	P48742	Q7Z5J1	Q9HA38
P29371	Q8IUH4	O15353	Q96S37	Q92613	P08253	Q9Y3P4
Q96A84	Q8N196	Q9UID3	P15509	Q8NHW3	Q6IPX1	Q6ZVY0
Q96HJ3	Q8WWB7	Q86XK2	Q7L1I2	Q7Z739	Q13330	P18054
Q86U06	P55085	Q96B70	Q96FZ7	Q8WVF1	Q96KM6	P07357
Q9BUQ8	Q9BWQ6	Q9Y2W1	P98187	Q9BYT3	P09630	Q9NR45
O75173	O15173	Q9BRL6	Q9H1P3	Q9NPF4	Q8N695	Q9Y6R0
Q6L8G4	Q96DH6	Q96Q80	Q15768	P32297	P35367	Q8WXH4
Q96CJ1	O43148	Q9H7C4	Q68DC2	Q9BQB4	Q6PF05	Q92839
Q06945	Q8N7B6	Q9NW68	O14771	Q8ND30	Q16204	Q494U1
Q8IYB3	Q96F46	P48547	P35908	Q5VV52	Q9BT76	Q6DD88
Q8TD17	Q9BSR8	Q8N8Q9	Q32MQ0	P78348	O15117	P27815
Q9Y608	P78364	Q8N6S4	Q14206	Q8IWW6	Q6ZSS7	Q9BXD5
Q01844	Q1KMD3	P42262	Q5VZV1	O95171	P30838	Q8NBT3
Q5T749	Q14289	Q92804	P30510	Q9BSK1	Q9UBC0	P49190
Q9BQI4	Q8WTR8	Q5TD94	Q9BY14	O75973	Q9NPH3	Q99435
Q5U5Q3	P23942	Q155Q3	Q9UGV2	Q15846	Q9NPY3	P08559
Q9UBS0	Q92570	Q9UM63	O60870	P35243	Q8IZ81	Q5VST6
Q8WXA9	Q8N344	Q8IZF4	Q96SU4	O95475	Q0D2K5	Q05682
Q96N64	O43474	Q06828	P58401	P49821	P30304	O43734
Q9Y215	Q9NVV0	Q86YV0	P83369	Q9Y644	P36871	P52565
Q8IY67	Q9HC73	Q13585	O00444	O75665	Q86VE9	P78424
Q8TD94	Q99523	O75636	P78386	A8MT70	O14718	Q96PD5
Q6ICG8	Q96G01	Q8NFU1	Q9H2S1	Q13470	P35221	Q96E09
Q5VUJ6	A8MYB1	O14910	Q5VTE6	Q96JG8	Q9Y3Q8	P62140
O43670	P49418	Q6ZNA4	Q9BVT8	Q9HCM4	Q9BXW6	Q9HDB5
O75690	O14656	Q96EI5	P57682	Q6ZRP7	Q8IWL2	P51168
Q92733	Q08397	Q10586	Q86XQ3	Q9NR20	P50749	Q5TFQ8
A8MVS5	Q9NX47	Q53GL0	Q9BZJ7	Q13616	A8MVW0	Q8TCW9
O75909	Q96K78	Q8NFZ4	O60563	A6PVC2	O15244	Q9GZZ6

O60422	Q13454	P43363	Q5T2Q4	Q9Y5K5	O60831	Q96ES6
Q9UNK4	Q8IU85	Q96IZ0	Q7RTT6	Q9UJ55	O15204	Q9BX46
Q9NWK9	P35250	Q9H501	Q9Y4X0	O94992	Q96AQ6	P15036
Q7Z5L4	Q15172	Q8NDN9	Q9Y2U5	Q9HB63	Q9NQI0	Q17R60
O94769	Q96K58	Q08050	Q9H489	Q92698	P23497	Q16611
Q13508	P51965	Q68DK7	O15400	O14829	Q8TDQ0	Q96E52
Q9HCG8	Q86W54	Q96M56	P35611	Q96DF8	Q96NZ8	P57075
Q701N2	Q9Y3X0	Q8N9I9	Q9UPG8	P57058	P15813	A6NI87
Q9UKR8	P28698	P20073	O95447	O15457	Q9NSE2	O95528
Q9UBC2	P58340	Q9NZU5	O15347	Q9HBU1	P43362	Q99467
Q6K0P9	Q9UBN1	Q8IVM8	Q9Y3E1	Q8NAT2	Q96PL5	Q9HCM7
O14511	Q96G74	Q8NI51	Q8N0Y7	P15848	Q6NUN9	Q8IZP7
P24592	P04259	Q8N441	Q15822	Q6VN20	Q16676	Q13432
Q9UBK2	P03372	Q9BQI9	Q9UPY6	A2RU54	Q9H0L4	O14983
O43312	Q04724	Q9HCJ2	A6NDD5	Q14512	Q96CC6	Q9BWH2
Q9H992	A8MTJ6	Q07001	Q93045	Q8N9T8	P02679	P01100
Q8IVE0	Q8NEE8	Q9H426	Q2WGJ6	P08779	Q9NQ36	O15287
A6NFI3	P57086	Q8N488	Q13490	Q6PI48	A6NEQ2	P61550
P49747	Q96S59	P35637	Q8IVF1	O95665	P49736	Q6ZMB5
Q9P0W5	O14798	P43681	P43351	Q9Y573	Q99439	Q5T8D3
O15522	Q86X45	P07225	Q86VE0	O60716	Q6FI81	Q00887
Q13495	P35663	O15354	P14866	Q9P0J7	Q8TF61	Q14582
P12110	Q96D42	Q15198	O95677	Q5T7P8	Q9NX00	O60447
Q9UJU5	Q96L92	Q5HY92	Q9NPF5	Q9P270	Q9P2H5	Q9HC62
Q9BRL7	O95857	Q9NV70	Q86TU7	Q86WP2	P13797	Q17RG1
Q02040	Q8IYE0	Q9NRU3	O76062	Q86U17	O14818	P11464
Q96A83	Q6ZN55	Q8TCQ1	Q8IXT2	Q9Y698	Q96GW7	Q9UKY1
Q96HR8	Q99933	Q9Y385	Q05195	Q5SZQ8	Q9C0D7	P07332
A1YPR0	A6NI28	Q2M3C6	Q9HBI6	O43424	P17858	Q9BZG8
O43823	Q14469	P02745	P34981	P19338	Q5U623	Q9NVM6
Q9BV99	Q99502	Q96C00	Q15427	Q9UPZ9	Q92543	Q9NX95
Q14249	Q92838	Q15003	P42263	Q9NWF9	O75593	Q06710
Q8IZU1	Q14162	Q969K3	Q9H2U9	Q5VSJ3	P22792	P56706
P08621	P50135	Q96M69	Q05215	O95793	P51811	O43602
Q6ZN18	P22626	Q7Z601	P21399	Q9UDW3	P08195	P81133
A6NKG2	Q9NR12	Q2VWA4	Q8N5D6	Q8IXK0	O43900	Q8WXT5
P48764	O00744	Q9Y6E2	Q9Y3I1	Q96FZ2	P42566	P15498
Q9UBY0	P41595	O14830	Q92597	O00470	Q86UQ0	P32970
Q6S5H4	Q969Z4	Q15389	P10645	O75582	Q6Q788	Q9UJX6
P01130	Q9H7X3	Q8N7S2	Q96T25	Q14254	P01011	P28358
Q13247	Q6NX45	Q9H8X9	Q3SYB3	Q6DHV7	Q17RD7	P49917
Q03692	Q9BPX3	Q15414	O00327	P24864	Q8TD46	Q5T953
Q9H2X3	Q8IUD6	Q0P6D6	Q16584	Q9UN37	P0C7T3	P09936
P98168	Q8NBS3	Q8WYN3	Q8IV90	O95630	Q06481	Q8IV53
Q66K66	Q8IY50	Q9BUX1	A1A4G5	Q6T423	P26440	P78410
P07196	Q12967	P49759	Q05932	Q6VB85	Q96EE3	Q6UB35
Q9NPG1	Q02161	Q7Z403	Q96J87	Q8WW52	P04798	Q5XUX0
P49959	Q5BKX6	Q9NRS6	Q9NWS9	A6NL05	Q9BY64	P23219
A6NEE1	P11473	Q8N609	P00747	P07358	Q5TC12	P27037
Q9BXX1	P19397	P20264	Q8N8X6	Q00587	Q86UN3	Q86VX9
P27658	P10909	Q69YI7	Q86VR2	P48167	Q9NV56	Q5TAA0
Q99684	P21810	Q16666	O95274	Q5VWJ9	Q14566	O60729
Q9Y6U7	Q9C009	Q96II8	P50914	O75771	Q8N109	Q6DHY5
Q9Y6P5	Q15485	Q9Y248	Q86VD1	Q5TC63	P10253	P14316

Q8TBJ5	Q8IV63	Q02252	P55072	Q8N9L1	Q8TC41	Q9BRP7
Q9NRA8	Q9Y2Y9	Q6IWH7	Q8NCF0	Q8IUE6	Q6IC10	Q86TL0
O94768	Q7Z4T9	Q9Y2T3	Q12951	A6NMB1	Q9H2H9	P17275
O75444	Q96EK4	Q86Y01	Q8IXQ5	Q9NP71	Q01105	Q8TAT5
A6NGB9	Q3KNW1	Q9H1B7	Q9HAS0	Q9BZG2	Q08477	Q9UGV6
Q49MG5	Q7Z794	Q8TDR0	Q86Y39	P98066	Q86TN4	Q9Y3Z3
Q9NP61	Q9NYF8	A6NNM8	Q8NGA6	O60602	O95716	O43482
Q00169	Q9NP31	Q6ZMH5	O14523	O00567	Q7Z713	Q8TD91
Q7Z6M4	Q9NYL2	A4IF30	Q8N6W0	P20592	Q14140	P98177
Q9C056	Q9UKA8	Q92917	Q96B86	Q14242	Q6UXY8	P11216
Q96S94	P49746	P16422	P35527	P08670	Q5T7W7	Q8ND76
O95343	O15381	Q7RTS5	O76083	Q8NB59	Q00056	P41219
P08588	Q9H4I2	Q9ULZ9	P31152	Q765P7	Q96C12	Q96LX7
Q9NYQ3	Q6AZY7	Q9UBG3	Q9BR39	Q8IYP9	Q9BXS4	P46059
Q7L1W4	Q8N2G8	O00258	Q5JT82	O75177	P87893	Q8NI36
Q96MW1	Q6ZMQ3	Q92688	O43825	Q96A25	P48023	Q5FWE3
Q9H3H9	Q9NTZ6	Q13042	Q99640	P57740	Q04756	Q9UK76
Q99750	Q8N912	Q9H5L6	Q70Z44	P09110	A6NEY8	Q86YP4
O60759	Q9Y4X1	Q7Z5L3	Q9Y6F7	Q5T230	Q05516	Q96KE9
Q96LX8	Q9BQ04	Q13148	Q9C0C4	P11171	Q9Y2D1	O95407
Q8IW35	Q9H7D7	Q96F86	Q96A33	Q9H2F9	Q99943	Q96DZ1
Q96QC0	Q7L1Q6	P35348	O75460	Q6UXK5	P43155	P26022
P14555	P00742	Q9BXJ7	Q9GZY0	Q8N8S7	Q9H6Y7	Q9UNY5
Q52LD8	Q8IVP5	Q9H6L5	Q14677	Q5W5X9	Q8IWE5	P59190
Q6UXH8	Q4G0S7	Q5EBM4	Q9BSF8	P54136	P30154	Q9UJV9
Q5T6S3	Q9NZV7	P52797	P01137	Q9H706	Q6ZN11	P14921
P09016	Q96JN0	O15226	Q9UKG4	Q9UGT4	Q99558	O15150
P01034	P19484	Q9H8N7	Q969E8	O00311	Q71H61	Q9BZL6
P28908	P35548	Q16656	O75069	Q9P2F5	Q05586	Q92806
Q8IWZ5	O75897	Q6ZVX7	Q70IA6	P49450	Q8IZJ4	Q96J84
Q9UBP4	Q96NY9	A6NDE4	Q8IWC1	Q9Y2G4	Q9NQC7	Q9UJT2
Q8N684	Q6PCD5	P78325	P33991	Q99942	O94806	O15516
Q15637	Q66K79	P01588	Q9UKQ2	Q9NZ72	Q17RR3	Q13400
Q8NCB2	Q9BT92	P05187	Q96MT3	Q495X7	O95926	Q8WXF7
Q6UX46	Q96JP5	Q8IWB1	P56696	Q16549	P25963	Q8TF66
P26927	Q16739	Q0P6D2	Q14934	Q9NSD7	P26442	O43464
Q9BZE0	Q9NYZ3	B1AKI9	Q9Y6X8	Q9NQ76	Q8TAM1	Q9Y4K0
Q8WUF5	P19544	Q96HL8	Q86TJ5	Q8WU20	Q9UBN4	Q99835
Q5H9T9	Q86YJ7	Q14123	Q96FV3	Q13263	Q9UBX2	P09871
P52756	P23515	Q96H86	Q9NXI6	Q9Y5U2	Q9NP95	Q8IYD2
Q8N2M8	Q8TF76	Q9BQG1	Q9H3T3	Q9Y4X5	Q7Z569	Q14CN4
O95072	Q8TDC3	Q96AE4	Q8WVD5	P78334	P21709	Q969W1
Q15424	Q99807	Q5U651	Q7Z7K2	Q96EZ4	Q9BTA9	Q96J42
Q96CX6	P20823	A0PJE2	P14651	Q9NVU0	Q8NGJ2	P05156
Q9Y5W3	Q86V86	P54845	Q6P4E1	P22735	Q9HC38	P26599
P11831	P37288	Q9UNW9	Q2TAY7	Q9UIC8	Q96PQ6	Q12874
O43525	P43366	Q8IXH6	Q8N0S6	A1E959	P54284	Q14145
O95157	Q8NFN8	P10643	Q9UH92	Q9H040	P16050	Q8TEU8
P26368	Q9UH99	Q16850	O95633	Q92871	O15131	Q07866
Q6P1X6	Q8N5H3	O15321	O95977	P13796	P00736	P19440
P49711	Q8N7X4	Q8IXF0	Q8TDD1	P08138	Q96DN2	P50336
P13647	Q86U44	Q8WUX9	Q14114	P78368	Q9BVQ7	Q9UBD0
Q7RTU3	P30559	Q14244	Q6UXH1	P28476	Q9H4M3	P48995
O15481	Q6ULP2	Q9UNF0	O75056	P43251	Q9BRQ8	Q6NUS8

P51815	Q9H607	O95196	Q8N446	Q8N8V2	A6NN73	O75427
P02746	Q8N5H7	Q0D2I5	A5D8T8	O75298	Q6ZNA5	O43435
Q8TF63	Q9UI46	Q86W10	P43243	P26367	Q6ZQW0	Q6IEG0
Q14940	Q4G112	Q9Y239	Q6NSX1	P20853	Q9H2L5	Q16348
Q9Y5Y6	P23246	Q9UNK9	P01233	Q96RN5	Q9ULR0	O15370
Q9NV58	Q9H0G5	P49674	Q8N807	P53801	Q13319	Q9UMS4
P50219	P04196	Q6A162	Q401N2	P51178	O95249	Q96I51
Q8N7H5	P36941	O75553	P47888	Q9GZV4	P16870	Q96MT7
P02748	Q5VUD6	Q96HB5	Q9H156	Q587J8	Q7Z628	Q9ULS6
Q4V9L6	Q9HCS2	Q86XJ1	Q15542	P14678	O60936	Q96RF0
P81274	Q9H4G1	Q5VT03	Q96M19	Q09161	P48637	Q9UHX1
P98155	Q9GZW5	P54792	Q5SWS2	Q8IUI8	P43353	Q6ZMY9
Q96KG9	Q96I99	Q8TCZ2	O00155	Q2TBC4	Q86TD4	P04908
P51991	Q9UIX4	O95076	Q01860	Q04726	Q9Y5Q9	Q13588
Q6P9F7	P11487	Q02779	Q9NR21	Q13077	P19622	Q9HCN8
O14977	Q8NA82	Q3LI81	Q6UXF7	P0C7X1	A1A4S6	O00451
P32456	Q9H5H4	Q13444	Q9UGM1	Q8IZP1	Q96N16	Q58WW2
Q9H091	P14780	Q6UX06	Q6ZNW5	O75400	O43365	Q9H313
Q6P2M8	B2RNN3	O15335	Q8NGN5	Q16671	O95292	Q7L099
Q16566	O43711	P0C7P1	Q9H2Y9	Q96FV2	Q6P444	P13866
Q9UQ88	Q68D42	O75364	Q8NHS0	O14672	P61328	Q6P2H8
Q6L8G8	P35218	P35222	Q9UMY4	Q8N7E2	Q9HCX4	Q8N787
Q5W0B1	Q8WWF5	P43354	Q6SZW1	Q13438	P25089	Q9UBE0
Q14246	Q5SRH9	Q8NFB5	Q9UJJ9	P29536	Q8NBT0	Q96JX3
P14222	Q9NRI5	Q75VX8	P52630	Q9Y330	P10826	Q9HC96
Q6ZMZ0	Q14541	Q9Y616	Q9Y5V3	P55884	Q9NR82	Q9Y3B6
Q8TD22	Q9Y2L9	Q96MG2	P59780	P22492	Q9HAP2	Q9H0H3
Q9UGJ0	P07359	Q9H330	P54840	Q8IZF7	O43399	Q9UFN0
Q9HC52	P05556	Q99527	Q8N693	Q684P5	Q9NQC8	P49763
P06733	Q969M3	P15408	P22888	Q96HE8	Q14002	Q96LR5
P52429	Q9UF56	Q6NUI6	A6NHJ4	Q53FD0	O95402	Q9H9A5
Q9UKJ1	Q8TB24	Q9UDV6	Q9GZV7	Q9Y210	Q3KR16	Q9NV06
Q8IZ63	Q13351	P48551	Q9BQQ7	O75541	Q6P3S1	P32926
Q6H9L7	P46109	Q9ULW3	P47813	Q14151	Q75WM6	P46459
Q99457	P21580	Q9UBY9	Q6JBY9	P0C7X2	Q9NWQ8	Q9HB29
Q92558	P00740	Q969I6	A6NF34	Q92945	O15480	Q11130
Q86YR6	Q9UBR4	Q9C073	P61011	Q96PK6	P01906	Q96MX0
P57764	Q9H228	P43364	P22736	Q16790	O60347	O75787
P31150	P11279	Q9ULX5	Q9P0W2	Q8N5A5	Q5VSL9	Q6DN90
Q8TBP5	Q8NE28	Q9UGQ3	P07204	Q13401	P51798	Q8WV24
P50395	Q15750	A1L4K1	O60224	P49640	P52952	Q8N6M5
Q9Y388	Q9H0B6	Q8N8Q8	Q96AX9	Q8IYL2	P15260	Q96LZ7
Q9BRN9	Q8IX06	Q5VZR2	Q13510	P14317	P0C5W0	Q9BYJ9
Q03395	Q5JTV8	P58304	Q96SL8	Q9H270	Q9NPA0	O14904
Q6DD87	O15075	P48380	Q15848	O00515	A6NHG9	Q13425
Q08722	Q92747	A9QM74	P22059	Q96A40	Q5VV42	Q9H2C8
Q6NUI2	Q8TDD2	Q8N7C7	Q9H7E9	Q8N4P6	Q96FE5	Q9Y6Y8
P13807	Q96ST8	Q9BTX7	Q96CM3	O43815	O43561	P15586
Q9BXU3	P19388	O00270	P51116	P85298	Q15835	Q9UIS9
Q6IPX3	Q9HBW1	P27918	Q8NFW5	Q6ZMT4	P01008	Q9Y2D0
Q15417	O00219	Q9UFB7	Q9BXP2	Q7Z5D8	Q15118	P04222
P21127	O95600	P18825	Q8WWY6	P0C1S8	Q96FK6	Q9Y5L0
Q8IYF3	Q9NQ90	Q8WUZ0	P47893	Q3ZCQ3	Q8IV04	Q16099
Q9Y285	P47211	A6NEQ0	P10914	Q29963	Q13144	P25774

P10155	Q12872	Q8N163	Q92504	Q96LZ2	Q6FI13	Q9H2K8
Q9NRY6	Q16820	Q9Y548	Q8TF05	Q07666	P32019	Q01831
O15079	Q8IWT1	Q9HAU4	Q6H8Q1	O14939	Q8NCA9	Q9BXJ8
Q96P44	O00421	Q06455	P33765	Q6UXT9	O15523	Q7Z7F0
P49756	P48739	Q9Y2C9	Q96MM6	Q8N7C4	Q5T0T0	Q8WY07
Q06587	Q8TCW7	P01871	Q8NA61	Q13368	Q9NPR2	Q13469
Q6ZNF1	Q03052	Q96EX2	Q99081	Q14442	Q96BR1	Q8IWF2
Q9UNA4	Q04671	Q8IX07	P48382	Q9H0C3	Q6NZI2	Q9BRS8
Q53HI1	O60359	Q9H1X3	Q9H8X2	Q8NHY2	P54725	A4D1S0
Q96DX8	Q9UQ84	Q9BZ81	Q96KK3	Q8NES3	Q8WX94	Q9BW04
P50454	Q6AI14	Q8WVK9	Q96HM7	P19438	P51858	Q9Y4R7
Q9H9R9	P42768	O14978	Q9H1E3	O75969	P54619	Q9NXF1
O15371	Q53ET0	Q7Z2K8	P41235	Q5T4W7	Q86TH1	O14994
P02747	Q96FM1	Q5SXM8	Q9Y2X9	Q15906	Q8N3J2	Q6P531
Q86WV6	Q99501	Q3MIW9	Q9UKN5	P13284	Q9P1T7	P49767
O14745	Q01546	Q6P597	A4D1F6	Q92826	Q8NH64	Q8N543
O00602	Q9H0U9	P16070	Q8IUX8	Q8TDU6	Q6NSZ9	P47775
Q8NBK3	P12081	Q9H560	Q92831	P10244	Q9NYY3	Q9UHR4
Q6ZS27	Q49AM1	Q6ZMY6	Q6PF06	P25942	Q5T310	P0C2W1
Q01130	Q9NPF0	P40692	P21757	Q9UI38	O75127	O15156
Q9BTE6	P18146	Q96GX1	Q71SY5	Q96DC7	Q13188	O43526
Q7L985	Q9H175	P23945	O43167	Q5SW24	Q14807	Q12778
Q5T700	Q9Y5L3	Q9BWF3	P59542	P56915	A1A4V9	P01584
P20333	Q9UPU6	Q96L46	Q96KN4	O60341	Q9Y3Q7	P46060
Q6ZNG2	Q8NEM7	O14512	O14863	P41214	P09429	Q9UPY5
Q07955	Q8WUX1	Q9BYV1	Q92932	Q15057	Q96JF0	Q8NEG0
O00401	Q5VWG9	Q96NE9	Q9H446	Q9H116	Q86VI1	O14669
Q9Y2T7	Q9H0F5	P35523	Q13277	Q9ULQ1	Q9UI14	Q7Z429
Q96T59	O14777	P59827	Q5TYM5	Q6P3W7	Q16401	P07741
Q13595	Q2QGD7	P55036	O95751	Q8N5Y2	Q9Y694	Q03167
A6NDB9	Q8N661	P30532	Q712K3	Q9NQZ2	Q8N187	Q9BZQ8
Q8TB68	Q9UBF1	Q496A3	Q8NC56	Q9H981	Q8N2U9	Q6UX07
Q9UGU5	Q9UDV7	Q96N87	Q6PII5	Q9Y6V7	Q8IU99	Q92966
P51114	Q7L0X0	Q96N20	Q8NE79	Q5U5R9	Q9BUZ4	P14384
O75175	Q8TBR4	Q9BV79	O00400	O75223	Q9Y4I5	Q96DI7
Q8N9Y4	Q99766	Q9UHX3	Q9P286	A6NER0	Q9HAT8	Q7Z6K1
Q96GP6	P10451	P15692	Q9C0B0	P25490	Q9H195	Q14416
P20226	P03951	Q9H3W5	Q13362	Q13761	Q92974	P50406
Q7Z4V5	Q9HCC6	Q8N831	Q8WU10	Q9H0M0	Q96HC4	O75459
O14638	P14210	Q5JU00	Q2TAC2	P23443	Q92796	Q7L7L0
O75864	Q8N1H7	P36639	Q86X60	P61978	P20749	O75901
Q5T8P6	Q96QU6	P54257	Q7L311	P78414	P05981	Q9H4T2
Q5H9R7	A6NK89	Q6PJF5	Q96HA8	P55107	Q8NC51	P29074
Q08117	Q9NXV6	Q9UL33	Q96RP8	P28566	Q8WYR1	Q53TS8
Q99613	Q6P9G9	A6NGH7	O15162	Q6PCB5	Q9NUT2	O95620
P05106	Q9BXQ6	Q8NGH8	Q9NRG4	Q9ULV4	Q9HBW9	Q13492
P32927	Q08345	O15350	P13671	Q9Y5S1	A0PJK1	Q15722
O43402	Q8TF30	A2A368	P80723	O14495	P55209	Q52WX2
Q969E4	Q8N8N0	O94927	Q9Y4L5	P22670	Q86XA0	P51159
Q8WUD6	P47895	Q8TC76	Q6P4H8	Q99853	Q13835	Q5U5Z8
Q86YR5	Q86YD5	Q9BTE7	Q8IYD1	Q96PV6	Q8N357	O14948
Q9P0L2	Q6ZMG9	Q02410	Q13064	Q9H6F5	Q9Y2T5	Q9H0U3
Q96N46	Q8WY54	Q14872	Q96D31	A0JLT2	O75509	O60861
Q9BZI1	Q15700	Q6ZW33	Q9NWS8	A6NED2	Q66LE6	Q76MJ5

Q9BU20	P78411	Q5MNZ6	Q14D04	Q9BW71	P23560	Q5TAQ9
P25067	P55318	Q68DQ2	P15884	Q96NL1	Q7Z6M3	Q9C0B5
Q8IUR6	Q8N5B7	Q7L211	Q9Y2B4	O14757	Q9BXJ0	Q9Y2B1
P43489	Q8IVQ6	Q9H4A6	Q96A26	Q5TBC7	Q92782	P49427
Q86VZ4	Q9BQ24	Q6ZNB6	P62906	Q9H4D5	Q5QJ38	O95935
O96009	O14908	P51449	Q8NEB7	Q9NQX3	Q9NYZ1	Q5TH74
P04156	Q9GZV8	Q9H7S9	Q9UBI9	Q14257	P17787	Q96JQ2
Q4VC44	Q7Z6B0	Q6ZQN5	Q9NQ55	Q8TC27	Q6L9T8	Q9HAW7
A6NM36	Q9H3Q1	P19634	P11230	Q86Y97	O43929	Q7L3T8
Q8IUK5	Q9UBH0	Q9NY99	P51957	P11362	Q9H2F5	Q99729
Q8IUG1	P0C7M4	Q9HB89	Q9BYV6	Q29960	Q9UK41	Q9Y2P5
Q9NW64	Q9Y5I7	Q8N2R0	Q9UNX3	Q9H4B7	Q96RT8	Q8NFK1
O43292	A8MXV6	Q9NNX6	Q8NDF8	Q9NUQ7	Q9UHK0	Q9H1B4
P82979	Q5SWH9	A6NNL0	Q92696	P85037	Q8TDI7	Q9ULT6
Q9P1Z0	A0AV02	Q9H611	Q6UUV7	Q8N752	Q9BQ87	Q8NE71
Q13435	Q8WVC0	Q96JB3	P15923	Q9BPX5	Q96LJ8	P49961
Q5H9L2	Q9UN36	Q5KU26	P18858	Q8N158	Q8IYB7	P63133
Q9ULC8	Q9NS66	O95948	Q7RTX9	Q96P15	O95888	P34947
A4D1S5	P13747	Q7Z353	Q9GZS9	Q8TCN5	Q18PE1	P21802
Q9BTK6	Q04727	Q86X95	A2RUB6	P11309	P49910	Q9H1Y3
Q96DV4	Q96S99	O15455	Q9UKR3	Q8IWL1	P02671	Q9Y230
Q15397	Q96CS7	Q16602	Q2VPK5	P56704	Q9NQ86	Q9BPX7
Q06413	P54727	Q9H305	Q9P209	P30499	Q13554	A0MZ66
P10827	Q9NX09	Q14764	P26012	O43422	Q13740	O95267
P58005	Q9Y5B6	O96018	P51606	Q641Q3	Q9H5I1	Q8NBR6
Q68DH5	P36021	A6ND36	Q96JB8	Q96HE9	P38432	Q9NXG6
Q5VYV0	A2VCL2	Q107X0	Q9Y6R9	Q15269	Q9HBX9	P21217
P35680	P02750	Q8NGJ6	Q9UK96	Q96N95	P56377	P43026
O94907	O43719	Q8TF74	P48039	P02686	Q9H169	Q8N5U6
Q8NGJ7	Q99504	Q86Y33	A6NLP5	P16444	Q5D1E8	Q8NAA4
O15144	Q9BRG2	P04198	Q6UX65	P35249	Q6NT55	Q13823
Q9H329	Q969G9	Q9Y5X5	P80370	Q96KA5	A4D2B8	A6NNZ2
Q8WWA1	P24941	Q9UH16	Q86Y38	Q9BXS9	P49642	Q9BXX5
Q96JW4	P57082	O14602	O95817	Q9UKT4	Q6NSJ5	Q14533
Q13507	Q9NYV9	Q9UNA1	O00418	Q6UX71	Q16643	Q9H1C4
Q8IWL3	Q8NFD2	Q06136	Q5T197	Q71RH2	Q96FB5	Q9UD71
Q9Y261	Q8NEK5	P18564	Q96NB2	Q2I0M4	P18428	Q96HH9
O00548	B1AL46	Q9BRJ9	Q9BSQ5	Q9Y572	Q96K37	P31273
P84157	P60660	Q9NXC2	O95865	A6NH17	Q16832	Q86VP3
Q96GC6	Q8IYT1	Q86YM7	Q9BY32	Q9H4H8	Q92667	Q16557
Q9BU76	Q9NXE8	Q5MNZ9	Q12765	Q9BQW3	Q969U6	Q9BXG8
Q8IXT5	P51164	Q16560	Q29865	O43493	Q9NNZ3	Q96T21
Q9BXU8	Q9ULK5	P10588	Q96QI5	Q9BRX9	P53597	Q8TBB1
Q6S545	Q16637	A6NMS3	Q8TDW5	Q13555	O43516	Q8NC54
Q8WUP2	Q6P4Q7	P04220	A6NGB7	Q96EZ8	Q8N4T0	Q9HBK9
Q9NUL5	P06576	Q02509	Q8N4L1	P49336	Q674R7	Q93070
A8MYZ6	Q969J5	Q13608	Q8N436	Q5RGM9	P49321	Q9H074
Q8TAK6	Q9H2X0	O15105	O96014	Q8NB90	Q14574	Q9HD64
Q8N8G2	O75953	P82912	P29401	Q8WXU2	P15169	Q4V328
Q92949	A6ND48	Q9C0H2	Q8N5C7	P54646	Q96E29	P78385
P02538	Q9BY78	P13945	P50052	O75715	Q8N2R8	Q9NVR2
Q12962	Q08499	P10074	P20674	Q16849	O43248	P51790
Q86UK7	Q9Y4A8	Q96P56	Q8N9U0	Q9NX18	A4GXA9	P32745
Q8WXF8	Q8NCK7	A6NHNO	Q7L5Y1	Q6TCH7	Q9UNN5	

Q53RY4	O14901	Q8IWX9	Q8IW70	Q9UJW0	P05091
P33121	O95980	Q16696	P03999	P16383	Q9UGI0
P49761	O15431	Q8IYU4	Q04725	Q8N584	Q92870
Q9H3T2	P52815	O14683	Q86U89	P20810	Q86Y46
Q96JE9	P21980	Q8N264	P84550	Q9NZR1	Q96PP9

Prilog 2

LISTA2: Proteini iz liste LISTA1, koji su anotirani GO terminom GO:0006355 – DNK-posredovana regulacija transkripcije. (Primena integrativnog algoritma za predikciju PPI na protein WT1)

Q6ZN18	O00470	Q08117	Q12951	Q9Y5V3	Q14151	Q96K62
O95076	Q86VE0	O00213	Q6ZQN5	Q8N344	P11831	Q9UFB7
Q15904	Q99750	Q92870	A8MTJ6	Q9BRJ9	Q8NEF9	P10074
Q9C0K0	Q9P1T7	Q8N187	Q92949	Q14872	P52630	O15156
P20749	O15226	P10275	P85037	Q13330	Q9H501	Q96C00
Q8N143	O15522	O00327	P58012	Q9UIS9	Q9NYF8	Q96SZ4
Q9HBU1	P52952	Q9UBL3	Q08050	Q8WWY6	Q01844	Q9H4T2
Q9NY43	Q9C056	Q96SQ7	O15353	P49736	Q14498	P10073
P49715	Q5T6S3	Q7L1Q6	Q12778	P33991	Q9UQ88	Q9UPU6
P49711	P14859	P35222	P98177	Q14566	O15156	P49910
Q8NI51	Q03052	O60716	O15409	P52815	Q04724	P17022
O75175	P20264	P07199	Q9C009	Q02779	P81133	O43670
Q53ET0	Q01860	Q9HC52	O43638	P50219	Q92966	Q9UDV6
Q6UUUV7	P78424	Q8N1H7	A8MYZ6	P35548	Q3KNW1	O14771
Q10586	Q9H4Q3	Q8TE02	Q2VWA4	P28698	Q9UNH7	Q9UNY5
Q92782	Q9GZV8	P16383	P29083	Q02078	Q9BXG8	O14978
Q9UHI6	Q9UKN5	Q9NV56	Q9Y692	Q06413	Q15637	Q96GC6
Q8TDD1	Q96IZ0	Q9Y5B6	Q05586	P54845	P23246	Q9Y2X9
Q8IXT2	Q8N7H5	Q8IUR6	P56915	Q8IXF0	Q02040	Q9UDV7
Q9UBC3	P51449	O15516	Q6ISB3	Q8IUM7	Q12872	Q96PQ6
Q9NPF5	P18615	Q86X95	Q99684	Q9Y4A8	Q9H5I1	Q96LW9
O75190	Q8N488	Q49AN0	Q14469	Q13469	Q86Y97	Q5BKZ1
A0AVK6	Q9BWF3	P24864	Q9HCC6	Q14934	O60224	Q9H8N7
P19419	Q9BQY4	O75909	Q4G112	P25963	Q7RTT6	Q96N95
Q96CJ1	P0C7M4	Q9UK58	Q9UBD0	Q9UHK0	Q7RTT4	Q8TD17
Q8N693	Q9Y230	Q96S94	P14317	Q9H6Q4	Q8N7U7	Q9H9D4
P53539	P82979	O60563	Q03014	Q9HCD5	Q9Y2W1	Q9NWS9
Q8TBJ5	P57086	P24941	P41235	Q9BQI9	P10827	Q6P9G9
Q96SL8	Q9GZW5	P49336	Q14541	P10588	O15455	Q7Z7K2
P15408	Q96KG9	P50750	P51858	P22736	Q9H3H9	Q6NSZ9
Q92908	Q14140	P49918	Q99729	P43354	Q969E4	O60304
Q86YP4	O95343	Q9H175	Q9BUJ2	Q92570	Q96E15	Q8TCN5
Q86WP2	Q8N196	Q8WYN3	O94992	Q13285	Q5H9L2	Q8NB15
Q9H116	O95475	Q8WXF8	Q9P0W2	Q16656	Q6IPX3	Q96KM6
Q9BZE0	O15105	Q6ZNG2	P09429	Q6ZNB6	Q99081	Q6ZMY9
Q8NEA6	Q9HAU4	Q8NFW5	Q7Z353	Q13516	P15923	Q5EBM4
A2RU54	P23497	Q9UBX2	Q9BQA5	Q8TAK6	P15884	Q8NEK5
O43719	O15370	Q9NQL9	O43365	Q7RTU3	Q9HCS4	Q86TJ5
Q9UGU5	O94993	Q9BQW3	Q00056	Q9UBC0	P19484	Q6ZN55
P20823	Q06945	P18146	Q92826	O95948	O14948	Q9BSK1
P35680	Q9UHI3	P11161	P14651	O60422	Q9UL49	Q86UQ0
Q92831	Q5VUG0	Q05215	O43248	P26367	Q9BQ87	Q96LX8
Q07666	Q8TAD8	O95905	Q7Z5D8	Q06710	Q04725	Q5T7W0
O75525	Q02446	Q92838	P09630	Q99453	Q04726	Q96CK0
Q92945	Q8TDD2	O75460	P31273	O75364	Q04727	Q6ZS27

Q13351	Q01892	Q76MJ5	P28358	Q9NZV7	P01137	Q96K58
O14901	O43435	O60869	P31277	P30048	Q03167	P0C7X2
Q9Y2Y9	O95935	P19622	P09016	Q9UBK2	Q13263	Q9H7X3
Q8TD94	P57082	Q9H2F5	P13378	P11309	P19438	Q9H7S9
Q9UIH9	P31314	Q52LR7	Q96JB3	Q9H307	P43489	Q6NUN9
Q9BXX1	O43763	P06733	P48551	Q9UM63	Q9H3D4	P51815
Q5JT82	O43711	P01588	P10914	Q9UPG8	O15350	Q96N20
Q9Y5W3	Q12962	P03372	P14316	Q9UHX1	Q5T230	Q96H86
P57682	Q5VWG9	Q9UL18	Q13568	O60828	P17480	Q9H5H4
O43474	Q15542	P49640	Q16666	Q6NZI2	P14921	Q7L3S4
O95600	Q9Y6J9	Q99502	P01584	Q13401	P15036	Q6NX45
Q8N5B7	Q5H9L4	O00167	Q9Y616	Q96AQ6	P01100	Q8NCA9
Q6ZMG9	Q13148	Q99504	P78414	Q9NVM4	O75444	Q6DD87
P84550	P20226	O95677	Q9BZI1	Q8WUF5	Q8NHW3	Q6ZN11
Q9NZU5	Q96EK4	Q96AE4	P78411	Q8NDN9	P10244	O75541
P48742	Q6Q0C0	Q9Y261	Q6NYC1	P22670	P04198	A6NHJ4
Q9UBR4	Q9Y3Q8	P55318	P17275	P48380	Q15906	Q96JP5
Q8WVC0	Q9H0M0	Q99853	Q14145	P48382	P15692	Q8IX07
Q99583	P19544	Q5VYV0	Q32MZ4	P10826	Q5SQQ9	Q8N5A5
Q05195	O00401	Q16676	O75427	O75582	Q8N8G2	P98168
Q14582	Q9Y2T7	Q9UJU5	Q96JN0	Q06587	P58304	P98169
Q9UH92	P25490	Q8WXT5	O60341	Q8N5U6	P11473	Q9UKY1
Q9HAP2	Q2QGD7	Q6VB85	Q8N5Y2	Q8WVD5	P04628	Q9Y6X8
Q9NP71	Q6H8Q1	Q5VV16	Q13495	Q06455	Q9Y330	
Q49AM1	Q9Y2D1	Q3SYB3	Q96RN5	Q13950	Q05516	
P56270	P18850	Q12946	A0JLT2	Q13761	O43167	
O95251	Q5U623	P55316	Q71SY5	P57059	Q9Y2Y4	
Q6ZN18	Q9ULW3	O75593	O95402	Q15424	Q9P1Z0	

Prilog 3

Lista proteina sa F(0,152) kao jednom od dve dominantne frekvence u informacionom spektru. (Funkcionalna karakterizacija domena proteina NISCH)

1A66_HUMAN	CRLS1_HUMAN	ISCA1_HUMAN	PHOS_HUMAN	SUGT1_HUMAN
1A68_HUMAN	CRML_HUMAN	K0408_HUMAN	PIEZ1_HUMAN	SUV92_HUMAN
ABCA2_HUMAN	CRP_HUMAN	K0754_HUMAN	PIGO_HUMAN	SYNC_HUMAN
ABCAC_HUMAN	CS073_HUMAN	K0895_HUMAN	PIGY_HUMAN	SYVN1_HUMAN
ABI2_HUMAN	CU037_HUMAN	K1467_HUMAN	PKHA9_HUMAN	SYWM_HUMAN
ABLM1_HUMAN	CU113_HUMAN	K1551_HUMAN	PKHM1_HUMAN	T185A_HUMAN
ACADM_HUMAN	CV026_HUMAN	K2C8_HUMAN	PLCB2_HUMAN	T185B_HUMAN
ACLY_HUMAN	CX024_HUMAN	KCNF1_HUMAN	PLCB4_HUMAN	T2R45_HUMAN
ACM2_HUMAN	CXCL3_HUMAN	KCNK2_HUMAN	PLCE1_HUMAN	TA2R_HUMAN
ACM4_HUMAN	CXCL5_HUMAN	KCNK3_HUMAN	PLSL_HUMAN	TAAR2_HUMAN
ACPL2_HUMAN	CYTS_HUMAN	KCP_HUMAN	PP4R1_HUMAN	TAAR8_HUMAN
ACPM_HUMAN	CYYR1_HUMAN	KDM6B_HUMAN	PP5D1_HUMAN	TAB1_HUMAN
ADA1B_HUMAN	D104A_HUMAN	KIF19_HUMAN	PPAC2_HUMAN	TAF11_HUMAN
ADAM5_HUMAN	D108A_HUMAN	KIF2B_HUMAN	PPAC_HUMAN	TAF12_HUMAN
ADIP_HUMAN	D108B_HUMAN	KISHA_HUMAN	PPM1L_HUMAN	TAM41_HUMAN
ADM2_HUMAN	D108C_HUMAN	KKCC1_HUMAN	PPM1M_HUMAN	TAOK1_HUMAN
ADPGK_HUMAN	DCAF7_HUMAN	KLC1_HUMAN	PPR18_HUMAN	TARSH_HUMAN
AKAP2_HUMAN	DCAKD_HUMAN	KPCL_HUMAN	PRA12_HUMAN	TASP1_HUMAN
AKD1A_HUMAN	DCE1_HUMAN	KR194_HUMAN	PRC2A_HUMAN	TBC24_HUMAN
AKND1_HUMAN	DDAH1_HUMAN	KV302_HUMAN	PRDX1_HUMAN	TBCC_HUMAN
ALKB1_HUMAN	DDX12_HUMAN	KV304_HUMAN	PRNT_HUMAN	TCF21_HUMAN
AMERL_HUMAN	DFFB_HUMAN	KV305_HUMAN	PRR18_HUMAN	TCL1B_HUMAN
AMRP_HUMAN	DGKG_HUMAN	KV310_HUMAN	PRR22_HUMAN	TDGF1_HUMAN
ANK1_HUMAN	DNAI1_HUMAN	KV312_HUMAN	PRRT3_HUMAN	TDGF3_HUMAN
ANKZ1_HUMAN	DNJC5_HUMAN	KV313_HUMAN	PRS46_HUMAN	TECTA_HUMAN
AP2A_HUMAN	DOCK3_HUMAN	LACRT_HUMAN	PS1C1_HUMAN	TEP1_HUMAN
AP2C_HUMAN	DOCK5_HUMAN	LARGE_HUMAN	PSA1_HUMAN	TF7L1_HUMAN
APBA1_HUMAN	DOPD_HUMAN	LCN8_HUMAN	PSG9_HUMAN	TF7L2_HUMAN
APT_HUMAN	DPP8_HUMAN	LDB3_HUMAN	PSMD2_HUMAN	TFAM_HUMAN
AQP9_HUMAN	DR4L2_HUMAN	LDHC_HUMAN	PSPN_HUMAN	THOP1_HUMAN
ARHG1_HUMAN	DR9C7_HUMAN	LHX6_HUMAN	PTCA_HUMAN	TIGD7_HUMAN
ARK72_HUMAN	DYH12_HUMAN	LIME1_HUMAN	PTPRS_HUMAN	TIM23_HUMAN
ARNT2_HUMAN	E41L3_HUMAN	LIPP_HUMAN	PUS10_HUMAN	TLN1_HUMAN
ARRB2_HUMAN	EDEM1_HUMAN	LMBRL_HUMAN	QCR6_HUMAN	TM110_HUMAN
ASAS1_HUMAN	EDEM2_HUMAN	LRC42_HUMAN	RAB3C_HUMAN	TM156_HUMAN
ASB10_HUMAN	EEPD1_HUMAN	LRC71_HUMAN	RAB8A_HUMAN	TM192_HUMAN
ASB4_HUMAN	EGLN1_HUMAN	LRIT3_HUMAN	RAB8B_HUMAN	TMM33_HUMAN
AT8A1_HUMAN	EI2BB_HUMAN	LSM5_HUMAN	RADIL_HUMAN	TMPSC_HUMAN
ATF6A_HUMAN	ELAF_HUMAN	LYRM1_HUMAN	RASM_HUMAN	TNNI2_HUMAN
ATG10_HUMAN	ENAM_HUMAN	MA2B1_HUMAN	RBM19_HUMAN	TNPO3_HUMAN
ATG3_HUMAN	ENPP1_HUMAN	MANBA_HUMAN	RENBP_HUMAN	TPC1_HUMAN
ATP6_HUMAN	ENPP5_HUMAN	MCM3_HUMAN	RFESD_HUMAN	TPGS2_HUMAN
ATPF1_HUMAN	EPAG_HUMAN	MCM6_HUMAN	RGPD1_HUMAN	TPH2_HUMAN
AUNIP_HUMAN	EPC2_HUMAN	MESD1_HUMAN	RGPD2_HUMAN	TPPC4_HUMAN
BAK_HUMAN	ESAM_HUMAN	MF2NB_HUMAN	RGS12_HUMAN	TPPC5_HUMAN
BAMBI_HUMAN	ETUD1_HUMAN	MGMT_HUMAN	RGS18_HUMAN	TPST2_HUMAN

BANP_HUMAN	EXT1_HUMAN	MGN2_HUMAN	RGSL_HUMAN	TR43B_HUMAN
BATF2_HUMAN	EYA4_HUMAN	MGT5B_HUMAN	RHG31_HUMAN	TR49C_HUMAN
BCL2_HUMAN	F195A_HUMAN	MKL1_HUMAN	RHOJ_HUMAN	TRBC1_HUMAN
BGAT_HUMAN	F27E2_HUMAN	MKS3_HUMAN	RIR1_HUMAN	TRBC2_HUMAN
BKRB2_HUMAN	FAKD2_HUMAN	MLRA_HUMAN	RNF24_HUMAN	TRI43_HUMAN
BLM_HUMAN	FARP2_HUMAN	MMP27_HUMAN	RNS12_HUMAN	TRI49_HUMAN
BMX_HUMAN	FBW10_HUMAN	MP2K2_HUMAN	RPN1_HUMAN	TRI67_HUMAN
BOLL_HUMAN	FBX4_HUMAN	MROH6_HUMAN	RS27A_HUMAN	TRIM6_HUMAN
BORG2_HUMAN	FCRLB_HUMAN	MRP7_HUMAN	RS30_HUMAN	TRIM9_HUMAN
BRD8_HUMAN	FINC_HUMAN	MT21A_HUMAN	RS6_HUMAN	TRIX3_HUMAN
BT2A2_HUMAN	FJX1_HUMAN	MTA2_HUMAN	RT24_HUMAN	TRM2_HUMAN
BUB3_HUMAN	FL2D_HUMAN	MUC15_HUMAN	S14L3_HUMAN	TRNK1_HUMAN
C1QT1_HUMAN	FOLR2_HUMAN	MUC1_HUMAN	S22A7_HUMAN	TRPM2_HUMAN
CA101_HUMAN	FOLR3_HUMAN	MUCB_HUMAN	S27A6_HUMAN	TSACC_HUMAN
CA2D1_HUMAN	FUBP3_HUMAN	MY18A_HUMAN	S35A5_HUMAN	TSC1_HUMAN
CAD24_HUMAN	FUCO_HUMAN	MYOM1_HUMAN	S35F6_HUMAN	TSN1_HUMAN
CAN10_HUMAN	FXL12_HUMAN	MZB1_HUMAN	S3TC1_HUMAN	TSN32_HUMAN
CASB_HUMAN	FXL14_HUMAN	NACC2_HUMAN	SALL4_HUMAN	TSSC1_HUMAN
CBPE_HUMAN	FXL15_HUMAN	NANGN_HUMAN	SAM9L_HUMAN	TTC40_HUMAN
CBPO_HUMAN	FXL16_HUMAN	NBEL1_HUMAN	SASH3_HUMAN	TTC9B_HUMAN
CBWD6_HUMAN	FXR1_HUMAN	NC2A_HUMAN	SCML1_HUMAN	TTL11_HUMAN
CC020_HUMAN	GABT_HUMAN	NCAM1_HUMAN	SDC2_HUMAN	TX261_HUMAN
CC038_HUMAN	GBA3_HUMAN	NCF2_HUMAN	SERC4_HUMAN	TXD15_HUMAN
CC121_HUMAN	GBRB1_HUMAN	NCTR1_HUMAN	SERF2_HUMAN	U2QL1_HUMAN
CC172_HUMAN	GELS_HUMAN	NDKM_HUMAN	SESN1_HUMAN	U633B_HUMAN
CCD36_HUMAN	GFI1B_HUMAN	NEST_HUMAN	SETD4_HUMAN	U633C_HUMAN
CCD86_HUMAN	GFRA1_HUMAN	NET1_HUMAN	SFRP1_HUMAN	UBE3A_HUMAN
CCD97_HUMAN	GGA2_HUMAN	NISCH_HUMAN	SFTA3_HUMAN	UBL3_HUMAN
CCDB1_HUMAN	GL54D_HUMAN	NJMU_HUMAN	SGIP1_HUMAN	UBL4B_HUMAN
CCL23_HUMAN	GLCM1_HUMAN	NKAI4_HUMAN	SGK71_HUMAN	UBP36_HUMAN
CD109_HUMAN	GLT16_HUMAN	NOL9_HUMAN	SHRM2_HUMAN	UBR3_HUMAN
CD177_HUMAN	GNA13_HUMAN	NP1L1_HUMAN	SHSA9_HUMAN	UD2B7_HUMAN
CDC42_HUMAN	GNAT2_HUMAN	NPTX2_HUMAN	SIA4C_HUMAN	UD3A2_HUMAN
CDC6_HUMAN	GOG6B_HUMAN	NSUN3_HUMAN	SIA7C_HUMAN	UFM1_HUMAN
CDK5_HUMAN	GOG6D_HUMAN	NT5M_HUMAN	SIA8F_HUMAN	UNC5B_HUMAN
CDRT1_HUMAN	GP108_HUMAN	NU214_HUMAN	SIAH3_HUMAN	UST_HUMAN
CEBPZ_HUMAN	GP173_HUMAN	NU2M_HUMAN	SIAL_HUMAN	VEGFB_HUMAN
CETN3_HUMAN	GPER_HUMAN	NUPL1_HUMAN	SIM14_HUMAN	VEZF1_HUMAN
CF217_HUMAN	GPKOW_HUMAN	O14L1_HUMAN	SLAF1_HUMAN	WDR6_HUMAN
CG029_HUMAN	GPR61_HUMAN	O2T10_HUMAN	SMAD6_HUMAN	WDR89_HUMAN
CG073_HUMAN	GPR88_HUMAN	O51G1_HUMAN	SMD2_HUMAN	WFDC2_HUMAN
CHD7_HUMAN	GPSM1_HUMAN	O52E5_HUMAN	SMIM5_HUMAN	WFDC6_HUMAN
CHM2B_HUMAN	GPSM2_HUMAN	O52H1_HUMAN	SMKR1_HUMAN	WIPI3_HUMAN
CJ011_HUMAN	GREM2_HUMAN	O52K2_HUMAN	SNP25_HUMAN	WNT11_HUMAN
CJ085_HUMAN	GRM3_HUMAN	O56B1_HUMAN	SNX3_HUMAN	WNT3A_HUMAN
CJ128_HUMAN	GROA_HUMAN	OCSTP_HUMAN	SNX4_HUMAN	WNT3_HUMAN
CK065_HUMAN	H1T_HUMAN	ODC_HUMAN	SOS2_HUMAN	WNT4_HUMAN
CKAP4_HUMAN	HBP1_HUMAN	ODF3B_HUMAN	SOSSC_HUMAN	WSB2_HUMAN
CKS2_HUMAN	HDC_HUMAN	OR2T4_HUMAN	SPA2L_HUMAN	XKR8_HUMAN
CL070_HUMAN	HERC5_HUMAN	OR2T7_HUMAN	SPAT7_HUMAN	YA043_HUMAN
CL17A_HUMAN	HIG1A_HUMAN	OR6J1_HUMAN	SPCS_HUMAN	YB025_HUMAN
CLM9_HUMAN	HNF1A_HUMAN	OR8J1_HUMAN	SPHK1_HUMAN	YC014_HUMAN
CLNK_HUMAN	HNF4G_HUMAN	OR8J3_HUMAN	SPI2A_HUMAN	YG001_HUMAN
CNDD3_HUMAN	HS3S6_HUMAN	OSBL8_HUMAN	SPI2B_HUMAN	YK032_HUMAN

CNGA3_HUMAN	HS71L_HUMAN	OVGP1_HUMAN	SPI2_HUMAN	YK042_HUMAN
CNO6L_HUMAN	HSP76_HUMAN	P2RY2_HUMAN	SPR2G_HUMAN	YP010_HUMAN
CO057_HUMAN	HXD12_HUMAN	P2RY8_HUMAN	SPSY_HUMAN	YV004_HUMAN
COA4_HUMAN	ICAM1_HUMAN	PA24F_HUMAN	SPXN2_HUMAN	YX002_HUMAN
COLI_HUMAN	IER5_HUMAN	PA2GC_HUMAN	SPXN5_HUMAN	ZBT7B_HUMAN
CP343_HUMAN	IFG15_HUMAN	PABP5_HUMAN	SRGEF_HUMAN	ZBTB1_HUMAN
CP3A4_HUMAN	IGHM_HUMAN	PACRL_HUMAN	SRP54_HUMAN	ZDHC1_HUMAN
CP3A5_HUMAN	IL19_HUMAN	PCID2_HUMAN	SSX9_HUMAN	ZFAN3_HUMAN
CP3A7_HUMAN	INAR1_HUMAN	PEX1_HUMAN	ST5_HUMAN	ZN292_HUMAN
CR063_HUMAN	INO80_HUMAN	PEX5_HUMAN	STAR8_HUMAN	ZN861_HUMAN
1A66_HUMAN	IQCF3_HUMAN	PGFRL_HUMAN	SUFU_HUMAN	

Prilog 4

Lista proteina sa F(0,426) kao jednom od dve dominantne frekvence u informacionom spektru. (Funkcionalna karakterizacija domena proteina NISCH)

1B13_HUMAN	EAF6_HUMAN	LPP_HUMAN	S4A10_HUMAN	ZFP90_HUMAN
1B40_HUMAN	EHBP1_HUMAN	LRCH3_HUMAN	SALL1_HUMAN	ZGPAT_HUMAN
1B44_HUMAN	EIF1B_HUMAN	LRP1B_HUMAN	SAM12_HUMAN	ZIC5_HUMAN
1B45_HUMAN	EIF1_HUMAN	LTBP3_HUMAN	SAMD7_HUMAN	ZN100_HUMAN
1B47_HUMAN	ELOV2_HUMAN	LUM_HUMAN	SAP25_HUMAN	ZN107_HUMAN
1B49_HUMAN	EMC8_HUMAN	LZTL1_HUMAN	SCLT1_HUMAN	ZN117_HUMAN
1B50_HUMAN	EMIL3_HUMAN	MA2B2_HUMAN	SCNNB_HUMAN	ZN121_HUMAN
5HT2A_HUMAN	ENK2_HUMAN	MAD1_HUMAN	SCO2_HUMAN	ZN124_HUMAN
5HT3A_HUMAN	ENK3_HUMAN	MAGD4_HUMAN	SCOT2_HUMAN	ZN132_HUMAN
A16L2_HUMAN	ENK5_HUMAN	MAP4_HUMAN	SCP2D_HUMAN	ZN135_HUMAN
A4_HUMAN	ENK6_HUMAN	MARH6_HUMAN	SCRN3_HUMAN	ZN136_HUMAN
AAPK2_HUMAN	EST5A_HUMAN	MATK_HUMAN	SCRT2_HUMAN	ZN157_HUMAN
AASS_HUMAN	EXOC3_HUMAN	MCM8_HUMAN	SDCG3_HUMAN	ZN169_HUMAN
ABCA2_HUMAN	F126B_HUMAN	MDN1_HUMAN	SDK2_HUMAN	ZN175_HUMAN
ABC2_HUMAN	F204A_HUMAN	MEDAG_HUMAN	SETMR_HUMAN	ZN177_HUMAN
ABCG2_HUMAN	F208B_HUMAN	MEP1A_HUMAN	SFTA2_HUMAN	ZN180_HUMAN
ABHD8_HUMAN	F214B_HUMAN	MER34_HUMAN	SGCE_HUMAN	ZN184_HUMAN
ABHEB_HUMAN	F219B_HUMAN	METRL_HUMAN	SHLB2_HUMAN	ZN195_HUMAN
ACTBM_HUMAN	F263_HUMAN	MEX3B_HUMAN	SIAH2_HUMAN	ZN208_HUMAN
ACTB_HUMAN	FA49B_HUMAN	MGDP1_HUMAN	SIAT6_HUMAN	ZN222_HUMAN
ACTC_HUMAN	FA83B_HUMAN	MGRN1_HUMAN	SIM12_HUMAN	ZN223_HUMAN
ADA11_HUMAN	FABPL_HUMAN	MGT4A_HUMAN	SIRPD_HUMAN	ZN226_HUMAN
AFF2_HUMAN	FANCB_HUMAN	MIA2_HUMAN	SKA2_HUMAN	ZN230_HUMAN
AGPD1_HUMAN	FAN_HUMAN	MKLN1_HUMAN	SKOR1_HUMAN	ZN251_HUMAN
AIMP1_HUMAN	FARP1_HUMAN	MMP16_HUMAN	SL9A7_HUMAN	ZN253_HUMAN
AKAP8_HUMAN	FB5L3_HUMAN	MMP24_HUMAN	SMI19_HUMAN	ZN254_HUMAN
AKT1_HUMAN	FBX10_HUMAN	MPPE1_HUMAN	SMIM2_HUMAN	ZN256_HUMAN
AL9A1_HUMAN	FBX16_HUMAN	MRO2A_HUMAN	SMIM7_HUMAN	ZN257_HUMAN
ALKB3_HUMAN	FBX33_HUMAN	MSI2H_HUMAN	SNAI1_HUMAN	ZN260_HUMAN
ALX3_HUMAN	FBX39_HUMAN	MSLNL_HUMAN	SNUT1_HUMAN	ZN266_HUMAN
ANFB_HUMAN	FCRL2_HUMAN	MTMR2_HUMAN	SNX11_HUMAN	ZN268_HUMAN
ANKH1_HUMAN	FCRL4_HUMAN	MTND_HUMAN	SOAT1_HUMAN	ZN273_HUMAN
ANR27_HUMAN	FEZF1_HUMAN	MYCL1_HUMAN	SODC_HUMAN	ZN284_HUMAN
ANTR2_HUMAN	FFAR3_HUMAN	MYC_HUMAN	SORT_HUMAN	ZN311_HUMAN
ANXA2_HUMAN	FGFR3_HUMAN	MYOZ1_HUMAN	SPA24_HUMAN	ZN322_HUMAN
AP2E_HUMAN	FHAD1_HUMAN	MZF1_HUMAN	SPAG4_HUMAN	ZN333_HUMAN
APAF_HUMAN	FKBP3_HUMAN	NACC1_HUMAN	SPAS2_HUMAN	ZN337_HUMAN
AQ12A_HUMAN	FKRP_HUMAN	NARF_HUMAN	SPB10_HUMAN	ZN33A_HUMAN
AQ12B_HUMAN	FLRT2_HUMAN	NBAS_HUMAN	SPDE3_HUMAN	ZN33B_HUMAN
ARPI9_HUMAN	FMOD_HUMAN	NBN_HUMAN	SPERT_HUMAN	ZN343_HUMAN
ARRD5_HUMAN	FOXA1_HUMAN	NCDN_HUMAN	SPG7_HUMAN	ZN347_HUMAN
ARVC_HUMAN	FPR1_HUMAN	NCOA3_HUMAN	SPN1_HUMAN	ZN358_HUMAN
AT11C_HUMAN	FPR2_HUMAN	NEK4_HUMAN	SPNXB_HUMAN	ZN367_HUMAN
AT5G1_HUMAN	FPR3_HUMAN	NEMO_HUMAN	SPT22_HUMAN	ZN37A_HUMAN
AT5G3_HUMAN	FSHB_HUMAN	NF2IP_HUMAN	SPXN1_HUMAN	ZN404_HUMAN
ATAD1_HUMAN	FXL21_HUMAN	NFYB_HUMAN	SPXN5_HUMAN	ZN416_HUMAN

ATLA1_HUMAN	G0S2_HUMAN	NFYC_HUMAN	SPYA_HUMAN	ZN418_HUMAN
ATM_HUMAN	GA2L3_HUMAN	NISCH_HUMAN	SRBS1_HUMAN	ZN426_HUMAN
ATP8_HUMAN	GAK11_HUMAN	NKAPL_HUMAN	STABP_HUMAN	ZN429_HUMAN
AXA2L_HUMAN	GAK12_HUMAN	NKX63_HUMAN	STIM2_HUMAN	ZN430_HUMAN
B2CL1_HUMAN	GAK1_HUMAN	NLRC5_HUMAN	STRN4_HUMAN	ZN431_HUMAN
B3GL2_HUMAN	GAK3_HUMAN	NMNA1_HUMAN	STX6_HUMAN	ZN433_HUMAN
B3GT2_HUMAN	GAK4_HUMAN	NOX4_HUMAN	SUCO_HUMAN	ZN439_HUMAN
BAI1_HUMAN	GAK5_HUMAN	NPNT_HUMAN	SYEP_HUMAN	ZN441_HUMAN
BAKOR_HUMAN	GAK6_HUMAN	NPT2A_HUMAN	SYF1_HUMAN	ZN442_HUMAN
BARD1_HUMAN	GAR1_HUMAN	NPY6R_HUMAN	SYLC_HUMAN	ZN443_HUMAN
BAX_HUMAN	GAREM_HUMAN	NRBP2_HUMAN	SYTL5_HUMAN	ZN454_HUMAN
BBS2_HUMAN	GBG12_HUMAN	NRDC_HUMAN	T150B_HUMAN	ZN460_HUMAN
BBS5_HUMAN	GBG3_HUMAN	NRSN2_HUMAN	T2R40_HUMAN	ZN484_HUMAN
BCAT2_HUMAN	GBG5_HUMAN	NUA4L_HUMAN	TAF1_HUMAN	ZN486_HUMAN
BCL6_HUMAN	GBG7_HUMAN	NUP53_HUMAN	TAF7_HUMAN	ZN490_HUMAN
BEX3_HUMAN	GBRR1_HUMAN	O10G7_HUMAN	TATD1_HUMAN	ZN493_HUMAN
BL1S3_HUMAN	GDF9_HUMAN	O10J6_HUMAN	TBC16_HUMAN	ZN496_HUMAN
BLK_HUMAN	GG2L6_HUMAN	O10R2_HUMAN	TBD2B_HUMAN	ZN514_HUMAN
BNC2_HUMAN	GG6L3_HUMAN	O14J1_HUMAN	TBX20_HUMAN	ZN526_HUMAN
C102A_HUMAN	GG6L4_HUMAN	O2T35_HUMAN	TBX3_HUMAN	ZN527_HUMAN
CA191_HUMAN	GG6L5_HUMAN	O52K1_HUMAN	TCAM1_HUMAN	ZN540_HUMAN
CA5BL_HUMAN	GG8L2_HUMAN	O5AC1_HUMAN	TDR15_HUMAN	ZN546_HUMAN
CABIN_HUMAN	GILT_HUMAN	ODO2_HUMAN	TEF_HUMAN	ZN555_HUMAN
CALD1_HUMAN	GINM1_HUMAN	OR1L1_HUMAN	TF2H3_HUMAN	ZN559_HUMAN
CASC4_HUMAN	GLIS3_HUMAN	OR1N2_HUMAN	TGIF1_HUMAN	ZN560_HUMAN
CB070_HUMAN	GLSL_HUMAN	OR2K2_HUMAN	THRSP_HUMAN	ZN561_HUMAN
CBLB_HUMAN	GNAS2_HUMAN	OR2T2_HUMAN	THS7B_HUMAN	ZN563_HUMAN
CC138_HUMAN	GOG8H_HUMAN	OR5BC_HUMAN	TIGD4_HUMAN	ZN564_HUMAN
CC151_HUMAN	GOG8I_HUMAN	OR5J2_HUMAN	TIMP4_HUMAN	ZN565_HUMAN
CC173_HUMAN	GOG8K_HUMAN	OR8J2_HUMAN	TLE4_HUMAN	ZN569_HUMAN
CC88B_HUMAN	GOG8M_HUMAN	ORC2_HUMAN	TLR7_HUMAN	ZN570_HUMAN
CC90B_HUMAN	GOG8N_HUMAN	ORC4_HUMAN	TM11A_HUMAN	ZN571_HUMAN
CCBE1_HUMAN	GOG8O_HUMAN	OSBL6_HUMAN	TM131_HUMAN	ZN573_HUMAN
CCD39_HUMAN	GOGA3_HUMAN	OSBL7_HUMAN	TM171_HUMAN	ZN575_HUMAN
CCD47_HUMAN	GORAB_HUMAN	OSCP1_HUMAN	TM209_HUMAN	ZN577_HUMAN
CCHL_HUMAN	GOT1B_HUMAN	OTOG_HUMAN	TM219_HUMAN	ZN583_HUMAN
CCM2L_HUMAN	GP107_HUMAN	OVOL3_HUMAN	TM50A_HUMAN	ZN589_HUMAN
CD032_HUMAN	GP141_HUMAN	OZF_HUMAN	TM86A_HUMAN	ZN595_HUMAN
CD28_HUMAN	GP158_HUMAN	PAHX_HUMAN	TMC6_HUMAN	ZN596_HUMAN
CD36_HUMAN	GPR20_HUMAN	PANX2_HUMAN	TMEM9_HUMAN	ZN607_HUMAN
CDC7_HUMAN	GPR33_HUMAN	PAQR6_HUMAN	TMLH_HUMAN	ZN613_HUMAN
CEAM3_HUMAN	GPR42_HUMAN	PAX2_HUMAN	TMM71_HUMAN	ZN615_HUMAN
CEAS1_HUMAN	GPS2_HUMAN	PCBP2_HUMAN	TMM78_HUMAN	ZN623_HUMAN
CECR2_HUMAN	GRK4_HUMAN	PCDBG_HUMAN	TMOD1_HUMAN	ZN625_HUMAN
CEP78_HUMAN	GRM1C_HUMAN	PCH2_HUMAN	TMPS5_HUMAN	ZN626_HUMAN
CERKL_HUMAN	GSTA1_HUMAN	PCY1B_HUMAN	TMPS9_HUMAN	ZN627_HUMAN
CFAD_HUMAN	GYS1_HUMAN	PDE12_HUMAN	TMTC1_HUMAN	ZN639_HUMAN
CG073_HUMAN	HAS3_HUMAN	PDP1_HUMAN	TNF18_HUMAN	ZN668_HUMAN
CGL_HUMAN	HDGL1_HUMAN	PEBP4_HUMAN	TNFC_HUMAN	ZN669_HUMAN
CH004_HUMAN	HDHD2_HUMAN	PEX3_HUMAN	TNNI3_HUMAN	ZN670_HUMAN
CH014_HUMAN	HECA2_HUMAN	PGK2_HUMAN	TOP3B_HUMAN	ZN672_HUMAN
CH047_HUMAN	HEMGN_HUMAN	PGRP1_HUMAN	TPC12_HUMAN	ZN674_HUMAN
CHDH_HUMAN	HIC2_HUMAN	PH4H_HUMAN	TRAF6_HUMAN	ZN675_HUMAN
CHFR_HUMAN	HIG2A_HUMAN	PHYIP_HUMAN	TRBP2_HUMAN	ZN676_HUMAN

CI163_HUMAN	HIG2B_HUMAN	PICK1_HUMAN	TRFL_HUMAN	ZN678_HUMAN
CIB2_HUMAN	HIKES_HUMAN	PIPSL_HUMAN	TRI46_HUMAN	ZN680_HUMAN
CIB4_HUMAN	HINT1_HUMAN	PJA2_HUMAN	TRI65_HUMAN	ZN682_HUMAN
CL033_HUMAN	HKR1_HUMAN	PKD1_HUMAN	TRM7_HUMAN	ZN684_HUMAN
CL3L1_HUMAN	HMGA1_HUMAN	PKHG6_HUMAN	TRPA1_HUMAN	ZN688_HUMAN
CLCC1_HUMAN	HMGB2_HUMAN	PKHO1_HUMAN	TS101_HUMAN	ZN695_HUMAN
CLD17_HUMAN	HMN4_HUMAN	PLAG1_HUMAN	TSNA1_HUMAN	ZN699_HUMAN
CLD22_HUMAN	HS3S2_HUMAN	PLIN3_HUMAN	TTC13_HUMAN	ZN700_HUMAN
CLD23_HUMAN	I27RA_HUMAN	PLPP_HUMAN	TTC32_HUMAN	ZN708_HUMAN
CLGN_HUMAN	ICAM4_HUMAN	PNCB_HUMAN	TTMP_HUMAN	ZN709_HUMAN
CMIP_HUMAN	ICAM5_HUMAN	POC1A_HUMAN	TUFT1_HUMAN	ZN710_HUMAN
CN028_HUMAN	IDLC_HUMAN	PP1R8_HUMAN	TYDP2_HUMAN	ZN714_HUMAN
CN079_HUMAN	IFT57_HUMAN	PP2BC_HUMAN	U633A_HUMAN	ZN716_HUMAN
CN16B_HUMAN	IG2R_HUMAN	PPWD1_HUMAN	UBA5_HUMAN	ZN717_HUMAN
CNNM1_HUMAN	IGSF6_HUMAN	PRD13_HUMAN	UBE2U_HUMAN	ZN718_HUMAN
CNNM2_HUMAN	IL10_HUMAN	PRD15_HUMAN	UBP12_HUMAN	ZN724_HUMAN
CNNM3_HUMAN	IMMT_HUMAN	PROSC_HUMAN	UBP46_HUMAN	ZN726_HUMAN
CNTD2_HUMAN	IMPA3_HUMAN	PRPK_HUMAN	UCP3_HUMAN	ZN728_HUMAN
CNTN3_HUMAN	IMPG2_HUMAN	PRR18_HUMAN	UN5CL_HUMAN	ZN729_HUMAN
CNTN4_HUMAN	INADL_HUMAN	PRR21_HUMAN	UNK_HUMAN	ZN730_HUMAN
CO5_HUMAN	INPP_HUMAN	PRRP_HUMAN	UQCC_HUMAN	ZN737_HUMAN
COL12_HUMAN	IPO13_HUMAN	PRRT4_HUMAN	VP13D_HUMAN	ZN746_HUMAN
COM1_HUMAN	IPO8_HUMAN	PS1C2_HUMAN	VPK7_HUMAN	ZN764_HUMAN
COMD4_HUMAN	IPYR2_HUMAN	PSMD4_HUMAN	WASH4_HUMAN	ZN771_HUMAN
COX7R_HUMAN	IPYR_HUMAN	PSPB_HUMAN	WDR17_HUMAN	ZN775_HUMAN
COXM1_HUMAN	IRK8_HUMAN	PTGES_HUMAN	WDR65_HUMAN	ZN781_HUMAN
CP011_HUMAN	IRX3_HUMAN	PTGIS_HUMAN	WFDC3_HUMAN	ZN782_HUMAN
CP046_HUMAN	K0895_HUMAN	PTPM1_HUMAN	WNK1_HUMAN	ZN788_HUMAN
CP090_HUMAN	K1C12_HUMAN	PUS7L_HUMAN	WNT2B_HUMAN	ZN791_HUMAN
CP250_HUMAN	K1C18_HUMAN	PYY2_HUMAN	YAED1_HUMAN	ZN799_HUMAN
CP46A_HUMAN	K1C28_HUMAN	QCR2_HUMAN	YAP1_HUMAN	ZN805_HUMAN
CP4Z2_HUMAN	KAD4_HUMAN	QSOX1_HUMAN	YB039_HUMAN	ZN823_HUMAN
CP7B1_HUMAN	KAPCA_HUMAN	RAB19_HUMAN	YETS4_HUMAN	ZN833_HUMAN
CR3L1_HUMAN	KAPCB_HUMAN	RAB35_HUMAN	YG006_HUMAN	ZN835_HUMAN
CRBA4_HUMAN	KAPCG_HUMAN	RAB3C_HUMAN	YI017_HUMAN	ZN836_HUMAN
CREL2_HUMAN	KATL1_HUMAN	RAB3D_HUMAN	YI023_HUMAN	ZN841_HUMAN
CRIM1_HUMAN	KAZRN_HUMAN	RAB41_HUMAN	YS017_HUMAN	ZN844_HUMAN
CROCC_HUMAN	KCQ1D_HUMAN	RAB43_HUMAN	Z280B_HUMAN	ZN878_HUMAN
CRY2_HUMAN	KDM5C_HUMAN	RASK_HUMAN	Z324B_HUMAN	ZN883_HUMAN
CSAD_HUMAN	KIFA3_HUMAN	RB8NL_HUMAN	Z354A_HUMAN	ZN891_HUMAN
CST9L_HUMAN	KKCC2_HUMAN	RBM27_HUMAN	Z354B_HUMAN	ZNF12_HUMAN
CTXN1_HUMAN	KLHL8_HUMAN	RBM44_HUMAN	Z355P_HUMAN	ZNF14_HUMAN
CYB5B_HUMAN	KMCP1_HUMAN	RBM8A_HUMAN	Z3H7A_HUMAN	ZNF20_HUMAN
CYFP1_HUMAN	KPB1_HUMAN	RELN_HUMAN	Z780A_HUMAN	ZNF25_HUMAN
CYS1_HUMAN	KPCD2_HUMAN	RFIP1_HUMAN	Z780B_HUMAN	ZNF26_HUMAN
CYTB_HUMAN	KR192_HUMAN	RFIP5_HUMAN	ZAN_HUMAN	ZNF2_HUMAN
DAPLE_HUMAN	KR194_HUMAN	RFOX3_HUMAN	ZBT11_HUMAN	ZNF30_HUMAN
DAZ1_HUMAN	KR195_HUMAN	RGS17_HUMAN	ZBT14_HUMAN	ZNF41_HUMAN
DC1L1_HUMAN	KR197_HUMAN	RL28_HUMAN	ZBT18_HUMAN	ZNF43_HUMAN
DCE2_HUMAN	KR198_HUMAN	RL36X_HUMAN	ZBT22_HUMAN	ZNF57_HUMAN
DCTN1_HUMAN	KR201_HUMAN	RLP24_HUMAN	ZBT24_HUMAN	ZNF66_HUMAN
DIAP1_HUMAN	KT3K_HUMAN	RN157_HUMAN	ZBT46_HUMAN	ZNF71_HUMAN
DOCK3_HUMAN	KV106_HUMAN	RN167_HUMAN	ZBT47_HUMAN	ZNF73_HUMAN
DPOA2_HUMAN	LAMA2_HUMAN	RN3P2_HUMAN	ZBT48_HUMAN	ZNF77_HUMAN

DPOD3_HUMAN	LEKR1_HUMAN	ROA1_HUMAN	ZBT49_HUMAN	ZNF7_HUMAN
DRAM2_HUMAN	LIMK1_HUMAN	ROBO2_HUMAN	ZC4H2_HUMAN	ZNF85_HUMAN
DRAXI_HUMAN	LIN41_HUMAN	ROCK1_HUMAN	ZCCHV_HUMAN	ZNF90_HUMAN
DRP2_HUMAN	LIPT_HUMAN	RPP25_HUMAN	ZCRB1_HUMAN	ZNF91_HUMAN
DSG1_HUMAN	LKHA4_HUMAN	RRAGA_HUMAN	ZEB2_HUMAN	ZNF92_HUMAN
DUS12_HUMAN	LLR1_HUMAN	RSAD1_HUMAN	ZF69B_HUMAN	ZNF93_HUMAN
DYR1A_HUMAN	LOXL3_HUMAN	RUVB1_HUMAN	ZFP14_HUMAN	ZNF98_HUMAN
E2AK1_HUMAN	LPAR6_HUMAN	S10AD_HUMAN	ZFP37_HUMAN	ZNF99_HUMAN
E2F6_HUMAN	LPHN1_HUMAN	S27A5_HUMAN	ZFP41_HUMAN	ZNT4_HUMAN
IB13_HUMAN	LPIN3_HUMAN	S46A3_HUMAN	ZFP69_HUMAN	ZNT7_HUMAN

Prilog 5

Lista AAS u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2 i pregled literature.

Gen	AAS	SNP/MUT	Bolest	Literatura
ASXL1	S25L	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	M29V	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	Q89R	SNP		(Sherry i sar. 2001)
ASXL1	A95V	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	C111Y	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	M168L	SNP		(Sherry i sar. 2001)
ASXL1	I268F	SNP		(Sherry i sar. 2001)
ASXL1	R394H	SNP		(Bejar i sar. 2011)
ASXL1	R402Q	MUT	MDS	(Bejar i sar. 2011)
ASXL1	S412F	SNP		(Sherry i sar. 2001)
ASXL1	E477Q	SNP		(Sherry i sar. 2001)
ASXL1	R489G	SNP		(Sherry i sar. 2001; Bejar i sar. 2011)
ASXL1	I507N	SNP		(Sherry i sar. 2001)
ASXL1	R541W	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	I552V	SNP		(Sherry i sar. 2001)
ASXL1	A636V	SNP		(Schnittger i sar. 2013)
ASXL1	A637T	SNP		(Schnittger i sar. 2013)
ASXL1	G643V	SNP		(Sherry i sar. 2001)
ASXL1	G652S	SNP		(Sherry i sar. 2001; Schnittger i sar. 2013)
ASXL1	C687R	SNP		(Sherry i sar. 2001)
ASXL1	T688M	SNP		(Schnittger i sar. 2013)
ASXL1	R693Q	SNP		(Schnittger i sar. 2013)
ASXL1	G704R	SNP		(Schnittger i sar. 2013)
ASXL1	P722R	SNP		(Schnittger i sar. 2013)
ASXL1	D741V	SNP		(Sherry i sar. 2001)
ASXL1	G744R	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
ASXL1	V751I	SNP		(Sherry i sar. 2001)
ASXL1	P779L	SNP		(Sherry i sar. 2001)
ASXL1	E788G	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
ASXL1	D799Y	SNP		(Sherry i sar. 2001; Schnittger i sar. 2013)
ASXL1	T806A	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	P808H	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))

ASXL1	T822I	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
ASXL1	L823S	SNP		(Bejar i sar. 2011)
ASXL1	K838R	SNP		(Sherry i sar. 2001; Abdel-Wahab i sar. 2011 (a))
ASXL1	S846N	MUT	MDS, MDS/MPN	(Jankowska i sar. 2011; Sugimoto i sar. 2012)
ASXL1	E865K	SNP		(Sherry i sar. 2001; Schnittger i sar. 2013)
ASXL1	M875V	SNP		(Schnittger i sar. 2013)
ASXL1	D979G	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	L983R	SNP		(Sherry i sar. 2001)
ASXL1	T1010M	SNP		(Sherry i sar. 2001)
ASXL1	G1026E	SNP		(Sherry i sar. 2001)
ASXL1	E1033V	SNP		(Sugimoto i sar. 2010)
ASXL1	P1035S	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
ASXL1	V1046M	SNP		(Schnittger i sar. 2013)
ASXL1	R1051C	SNP		(Schnittger i sar. 2013)
ASXL1	P1089S	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
ASXL1	V1092M	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
ASXL1	D1127N	SNP		(Sherry i sar. 2001)
ASXL1	M1161V	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	S1166R	SNP		(Sherry i sar. 2001; Bejar i sar. 2011)
ASXL1	R1171S	SNP		(Brecqueville i sar. 2012)
ASXL1	G1187V	SNP		(Sugimoto i sar. 2010)
ASXL1	P1197S	SNP		(Schnittger i sar. 2013)
ASXL1	R1247C	SNP		(Sherry i sar. 2001)
ASXL1	M1249V	SNP		(Sherry i sar. 2001; Abdel-Wahab i sar. 2011 (a); Schnittger i sar. 2013)
ASXL1	D1252G	SNP		(Sherry i sar. 2001)
ASXL1	P1259L	SNP		(Sherry i sar. 2001)
ASXL1	V1297I	SNP		(Sherry i sar. 2001)
ASXL1	A1312V	SNP		(Sherry i sar. 2001; Bejar i sar. 2011; Schnittger i sar. 2013)
ASXL1	R1316H	SNP		(Schnittger i sar. 2013)
ASXL1	L1325F	SNP		(Sherry i sar. 2001; Martínez-Avilés i sar. 2012)
ASXL1	G1339E	MUT	MPN, karcinom dojke	(Boulwood i sar. 2010; Forbes i sar. 2011)
ASXL1	A1380V	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	L1395V	SNP		(Sherry i sar. 2001; Pérez i sar. 2011)

ASXL1	G1397S	SNP		2010) (Sherry i sar. 2001; Bejar i sar. 2011; Schnittger i sar. 2013)
ASXL1	M1402L	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	S1428P	SNP		(Sherry i sar. 2001; Schnittger i sar. 2013)
ASXL1	F1463L	SNP		(Schnittger i sar. 2013)
ASXL1	A1486T	MUT	MDS	(Bejar i sar. 2011)
ASXL1	T1494I	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	S1496N	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	S1497G	SNP		(Abdel-Wahab i sar. 2011 (a))
ASXL1	T1498M	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
ASXL1	I1514L	SNP		(Sherry i sar. 2001)
ASXL1	C1519Y	MUT	MPN ili MDS/MPN	(Abdel-Wahab i sar. 2011 (a))
EZH2	I55M	SNP		(Sherry i sar. 2001)
EZH2	M134K	MUT	MDS	(Bejar i sar. 2011)
EZH2	L149Q	MUT	MPN	(Forbes i sar. 2011; Guglielmelli i sar. 2011)
EZH2	D185H	SNP		(Sherry i sar. 2001)
EZH2	E249K	MUT	MDS	(Ernst i sar. 2010; Nikoloski i sar. 2010; Forbes i sar. 2011)
EZH2	Q250E	SNP		(Sherry i sar. 2001)
EZH2	N317S	SNP		(Sherry i sar. 2001)
EZH2	N361S	SNP		(Bejar i sar. 2011)
EZH2	A384T	MUT	MPN, karcinom pluća	(Ernst i sar. 2010; Forbes i sar. 2011)
EZH2	A482T	SNP		(Sherry i sar. 2001; Grossmann i sar. 2012)
EZH2	C534R	MUT	MDS	(Bejar i sar. 2011)
EZH2	T568I	MUT	MPN	(Forbes i sar. 2011; Guglielmelli i sar. 2011)
EZH2	L575P	MUT	MDS	(Bejar i sar. 2011)
EZH2	Y641C	MUT	AML, MDS, limfoidni maligniteti	(Morin i sar. 2010; Capello i sar. 2011; Forbes i sar. 2011; McCabe i sar. 2012)
EZH2	Q648E	MUT	MDS	(Bejar i sar. 2011)
EZH2	A651V	MUT	MDS	(Ernst i sar. 2010; Forbes i sar. 2011)
EZH2	D652H	MUT	MDS	(Bejar i sar. 2011; Wang i sar. 2013)
EZH2	D659A	MUT	AML	(Forbes i sar. 2011)
EZH2	F655L	MUT	MDS	(Bejar i sar. 2011)
EZH2	L669V	MUT	MDS	(Nikoloski i sar. 2010; Forbes i sar. 2011)

EZH2	D672H	MUT	MPN	(Forbes i sar. 2011; Guglielmelli i sar. 2011)
EZH2	V674M	MUT	MPN, AML	(Forbes i sar. 2011; Guglielmelli i sar. 2011)
EZH2	R679C	MUT	MPN,	(Ernst i sar. 2010; Forbes i sar. 2011)
EZH2	R685H	MUT	MDS/MPN MDS, MPN, MDS/MPN, AML	(Ernst i sar. 2010; Makishima i sar. 2010; Bejar i sar. 2011; Forbes i sar. 2011; Grossmann i sar. 2011; Guglielmelli i sar. 2011; Jankowska i sar. 2011)
EZH2	N688T	MUT	AML	(Makishima i sar. 2010; Forbes i sar. 2011)
DNMT3A	E30A	SNP		(Sherry i sar. 2001)
DNMT3A	G278S	SNP		(Sherry i sar. 2001)
DNMT3A	S352N	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	K468R	MUT	AML	(Ley i sar. 2010)
DNMT3A	V480G	SNP		(Sherry i sar. 2001)
DNMT3A	N501S	SNP		(Sherry i sar. 2001)
DNMT3A	Y533C	MUT	AML, karcinom pluća	(Forbes i sar. 2011)
DNMT3A	G543C	MUT	AML, MDS, limfoidni maligniteti	(Forbes i sar. 2011; Shen i sar. 2011; Yan i sar. 2011; Marková i sar. 2012; Grossmann i sar. 2013)
DNMT3A	W581C	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	R635W	MUT	MDS, AML	(Forbes i sar. 2011; Thol i sar. 2011 (a); Walter i sar. 2011; Marková i sar. 2012)
DNMT3A	V636M	MUT	AML	(Forbes i sar. 2011; Thol i sar. 2011 (a); Fried i sar. 2012)
DNMT3A	I655A	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	R676W	MUT	AML	(Forbes i sar. 2011)
DNMT3A	G699S	MUT	MDS	(Forbes i sar. 2011)
DNMT3A	S714C	MUT	MDS, AML, karcinom dojke	(Brecqueville i sar. 2011; Forbes i sar. 2011; Shen i sar. 2011; Thol i sar. 2011 (a); Walter i sar. 2011; Marková i sar. 2012)
DNMT3A	P718L	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	R729Q	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	R729W	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	F731C	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)

DNMT3A	Y735C	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	R736H	MUT	AML, limfoidni maligniteti	(Ley i sar. 2010; Forbes i sar. 2011; Grossmann i sar. 2013)
DNMT3A	L737R	MUT	AML	(Forbes i sar. 2011; Walter i sar. 2011)
DNMT3A	A741V	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	R749C	MUT	MPN, AML	(Forbes i sar. 2011; Brecqueville i sar. 2012)
DNMT3A	R749P	SNP		(Sherry i sar. 2001)
DNMT3A	F752L	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	W753G	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	L754R	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	S770W	MUT	AML	(Forbes i sar. 2011; Walter i sar. 2011)
DNMT3A	R771L	MUT	MDS	(Forbes i sar. 2011; Walter i sar. 2011)
DNMT3A	R771Q	MUT	MPN, AML, karcinom pluća	(Forbes i sar. 2011)
DNMT3A	S775P	SNP		(Sherry i sar. 2001)
DNMT3A	R792H	MUT	AML, limfoidni maligniteti, karcinom pluća	(Ley i sar. 2010; Forbes i sar. 2011; Grossmann i sar. 2013)
DNMT3A	W795C	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	R803S	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	K829R	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	K841Q	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	P849R	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	M880V	MUT	MPN	(Forbes i sar. 2011; Rao i sar. 2012)
DNMT3A	R882C	MUT	MDS, MPN, MDS/MPN, AML, limfoidni maligniteti	(Ley i sar. 2010; Yamashita i sar. 2010; Ewalt i sar. 2011; Forbes i sar. 2011; Jankowska i sar. 2011; LaRochelle i sar. 2011; Lin i sar. 2011; Shen i sar. 2011; Stegelmann i sar. 2011; Thol i sar. 2011 (a); (b); Yan i sar. 2011; Fried i sar. 2012; Li i sar. 2012; Marková i sar. 2012; Rao i sar. 2012; Sekeres i sar. 2012;

DNMT3A	R882H	MUT	MDS, MPN, MDS/MPN, AML, limfoidni maligniteti	Grossmann i sar. 2013; Lu i sar. 2013) (Ley i sar. 2010; Yamashita i sar. 2010; Abdel-Wahab i sar. 2011 (b); Ewalt i sar. 2011; Forbes i sar. 2011; Jankowska i sar. 2011; LaRoche i sar. 2011; Lin i sar. 2011; Shen i sar. 2011; Stegelmann i sar. 2011; Thol i sar. 2011 (a); (b); (c); Walter i sar. 2011; Wlodarski i sar. 2011; Yan i sar. 2011; Couronné i sar. 2012; Fried i sar. 2012; Li i sar. 2012; Marková i sar. 2012; Ottone i sar. 2012; Pezzi i sar. 2012; Sekeres i sar. 2012; Simon i sar. 2012; Grossmann i sar. 2013; Lu i sar. 2013; Shih i sar. 2013)
DNMT3A	R882P	MUT	MDS/MPN, AML	(Ley i sar. 2010; Forbes i sar. 2011; Jankowska i sar. 2011; LaRoche i sar. 2011; Lin i sar. 2011; Shen i sar. 2011; Marková i sar. 2012)
DNMT3A	L889P	MUT	AML	(Forbes i sar. 2011; Fried i sar. 2012)
DNMT3A	V897D	MUT	AML	(Forbes i sar. 2011; Shen i sar. 2011; Yan i sar. 2011)
DNMT3A	P904L	MUT	AML, limfoidni maligniteti	(Ley i sar. 2010; Forbes i sar. 2011; Walter i sar. 2011; Grossmann i sar. 2013)
DNMT3A	F909C	MUT	AML	(Ley i sar. 2010; Forbes i sar. 2011)
DNMT3A	A910V	MUT	AML, karcinom endometriuma, intestinalni karcinom	(Forbes i sar. 2011; Fried i sar. 2012)
TET2	H8D	SNP		(Sherry i sar. 2001)
TET2	P22L	MUT	AML	(Forbes i sar. 2011; Ottone i sar. 2012)
TET2	P29R	SNP		(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	L34F	SNP		(Sherry i sar. 2001; Abdel-Wahab i sar. 2009; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i

TET2	P101H	MUT	AML	sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	P174H	SNP		(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	A215V	SNP		(Sherry i sar. 2001; Abdel-Wahab i sar. 2009; Weissmann i sar. 2012)
TET2	V218M	SNP		(Smith i sar. 2010)
				(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	A241V	SNP		(Olcaydu i sar. 2011)
TET2	S282F	MUT	AML	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	A308T	MUT	AML	(Jankowska i sar. 2009)
TET2	N312S	MUT	AML	(Abdel-Wahab i sar. 2009; Forbes i sar. 2011)
TET2	L346P	MUT	MDS/MPN	(Forbes i sar. 2011)
TET2	E350K	MUT	AML	(Makishima i sar. 2011)
TET2	G355D	SNP		(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	P363L	SNP		(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	S393C	SNP		(Langemeijer i sar. 2011)
TET2	P399L	MUT	MDS/MPN	(Jankowska i sar. 2009)
TET2	G429R	SNP		(Sherry i sar. 2001)
TET2	S460F	MUT	MDS/MPN, AML	(Abdel-Wahab i sar. 2009; Langemeijer i sar. 2011)
TET2	P472A	SNP		(Sherry i sar. 2001)
TET2	P474L	SNP		(Sherry i sar. 2001)
TET2	P480T	SNP		(Smith i sar. 2010)
TET2	T492S	MUT	AML	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	Q548L	MUT	AML	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	D666G	MUT	MDS/MPN	(Abdel-Wahab i sar. 2009)
TET2	S716L	SNP		(Smith i sar. 2010)

TET2	P761L	SNP		(Sherry i sar. 2001)
TET2	N767D	SNP		(Sherry i sar. 2001; Metzeler i sar. 2011)
TET2	D768V	SNP		(Smith i sar. 2010)
TET2	G773V	MUT	AML	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	E788L	MUT	AML	(Forbes i sar. 2011; Nomdedéu i sar. 2012)
TET2	Q810R	SNP		(Sherry i sar. 2001; Martínez-Avilés i sar. 2012)
TET2	R814C	SNP		(Chou i sar. 2011)
TET2	S817T	MUT	MDS/MPN	(Jankowska i sar. 2009)
TET2	S826I	MUT	AML	(Forbes i sar. 2011; Nomdedéu i sar. 2012)
TET2	F868L	SNP		(Chou i sar. 2011)
TET2	A912G	SNP		(Sherry i sar. 2001)
TET2	H924R	SNP		(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010)
TET2	P941S	MUT	MDS/MPN	(Abdel-Wahab i sar. 2009)
TET2	P968R	SNP		(Langemeijer i sar. 2011)
TET2	E1010D	SNP		(Metzeler i sar. 2011)
TET2	S1015N	SNP		(Smith i sar. 2010)
TET2	I1025T	SNP		(Smith i sar. 2010)
TET2	S1039L	SNP		(Chou i sar. 2011)
TET2	E1073V	SNP		(Abdel-Wahab i sar. 2009)
TET2	Q1084P	SNP		(Sherry i sar. 2001; Abdel-Wahab i sar. 2009; Jankowska i sar. 2009; Langemeijer i sar. 2009; Nibourel i sar. 2010; Smith i sar. 2010; Weissmann i sar. 2012)
TET2	L1120M	SNP		(Sherry i sar. 2001)
TET2	C1135Y	MUT	MDS, MPN, MDS/MPN	(Abdel-Wahab i sar. 2009; Thiel i sar. 2011; Pérez i sar. 2012)
TET2	E1144A	MUT	MPN	(Forbes i sar. 2011)
TET2	R1167T	MUT	AML	(Jankowska i sar. 2009)
TET2	C1193W	MUT	MDS, MDS/MPN	(Abdel-Wahab i sar. 2009; Kosmider i sar. 2009 (b); Smith i sar. 2010)
TET2	K1197R	MUT	MDS	(Rocquain i sar. 2010; Forbes i sar. 2011)
TET2	V1199I	MUT	AML	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	S1204C	MUT	MPN	(Abdel-Wahab i sar. 2009)
TET2	L1210P	MUT	MPN	(Schaub i sar. 2010)
TET2	C1211Y	MUT	MDS/MPN,	(Kohlmann i sar. 2010; Forbes i sar.

			AML	2011; Kosmider i sar. 2011; Metzeler i sar. 2011)
TET2	R1214W	MUT	MDS, MDS/MPN, AML	(Abdel-Wahab i sar. 2009; Kosmider i sar. 2009 (b); Langemeijer i sar. 2009; Flach i sar. 2010; Weissmann i sar. 2012)
TET2	H1219Q	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	H1219R	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	E1222G	MUT	AML	(Abdel-Wahab i sar. 2010; Forbes i sar. 2011)
TET2	L1229R	MUT	AML	(Forbes i sar. 2011; Weissmann i sar. 2012)
TET2	E1234G	MUT	AML	(Abdel-Wahab i sar. 2010; Forbes i sar. 2011)
TET2	D1242V	MUT	MPN, AML	(Abdel-Wahab i sar. 2009; Tefferi i sar. 2009 (b); Beer i sar. 2010; Forbes i sar. 2011; Pronier i sar. 2011)
TET2	Y1245S	MUT	MPN	(Abdel-Wahab i sar. 2009)
TET2	L1248P	MUT	AML	(Forbes i sar. 2011; Weissmann i sar. 2012)
TET2	G1256A	MUT	AML	(Forbes i sar. 2011)
TET2	G1256D	MUT	AML	(Forbes i sar. 2011)
TET2	G1256R	MUT	AML	(Abdel-Wahab i sar. 2010)
TET2	R1261C	MUT	MPN, MDS/MPN, AML	(Abdel-Wahab i sar. 2009; Kohlmann i sar. 2010; Forbes i sar. 2011; Metzeler i sar. 2011; Moran-Crusio i sar. 2011; Weissmann i sar. 2012)
TET2	R1261H	MUT	MDS/MPN, AML	(Abdel-Wahab i sar. 2009; Kohlmann i sar. 2011; Gaidzik i sar. 2012)
TET2	C1263R	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	N1266Y	MUT	AML	(Forbes i sar. 2011)
TET2	T1270A	MUT	AML	(Forbes i sar. 2011; Nomdedéu i sar. 2012)
TET2	C1271S	MUT	AML	(Forbes i sar. 2011; Weissmann i sar. 2012)
TET2	C1271W	MUT	MDS, MDS/MPN, AML	(Nibourel i sar. 2010; Forbes i sar. 2011; Voso i sar. 2011; Yoshida i sar. 2011; Bacher i sar. 2012)
TET2	C1271Y	MUT	AML	(Forbes i sar. 2011; Weissmann i

TET2	G1275E	MUT	MDS, MPN, AML	sar. 2012) (Schaub i sar. 2010; Smith i sar. 2010; Chou i sar. 2011; Forbes i sar. 2011)
TET2	G1275R	MUT	MPN	(Forbes i sar. 2011; Olcaydu i sar. 2011)
TET2	G1282D	MUT	MDS, AML	(Nibourel i sar. 2010; Smith i sar. 2010; Forbes i sar. 2011)
TET2	S1286P	MUT	MDS	(Rocquain i sar. 2010)
TET2	F1287L	MUT	MDS/MPN	(Jankowska i sar. 2009)
TET2	G1288S	MUT	AML	(Chou i sar. 2011; Forbes i sar. 2011)
TET2	G1288V	MUT	MPN, MDS/MPN	(Abdel-Wahab i sar. 2009; Smith i sar. 2010)
TET2	C1289Y	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011; Weissmann i sar. 2012)
TET2	W1291R	MUT	MDS, AML	(Langemeijer i sar. 2009; Chou i sar. 2011; Forbes i sar. 2011)
TET2	S1292R	MUT	AML	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	C1298W	MUT	AML	(Jankowska i sar. 2011; Makishima i sar. 2011)
TET2	C1298Y	MUT	MDS/MPN, AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	K1299N	MUT	MDS/MPN	(Jankowska i sar. 2009)
TET2	R1302G	MUT	MPN, AML	(Delhommeau i sar. 2009; Jankowska i sar. 2009; Pronier i sar. 2011)
TET2	E1318G	MUT	MDS/MPN	(Jankowska i sar. 2009)
TET2	L1340R	MUT	AML	(Forbes i sar. 2011; Weissmann i sar. 2012)
TET2	A1355P	MUT	MDS	(Smith i sar. 2010)
TET2	R1359C	MUT	MPN, MDS/MPN, AML	(Abdel-Wahab i sar. 2009; Tefferi i sar. 2009 (a); (b); Abdel-Wahab i sar. 2010; Gelsi-Boyer i sar. 2010; Chou i sar. 2011; Forbes i sar. 2011)
TET2	R1359H	MUT	MDS/MPN, AML	(Kosmider i sar. 2009 (b); Chou i sar. 2011; Forbes i sar. 2011)
TET2	L1360Q	MUT	MDS	(Rocquain i sar. 2010)
TET2	G1361D	MUT	MDS/MPN, AML	(Kohlmann i sar. 2010; Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	G1361R	MUT	MPN	(Abdel-Wahab i sar. 2009)
TET2	R1366H	MUT	AML	(Couronné i sar. 2010; Rocquain i sar. 2010; Chou i sar. 2011; Metzeler i sar. 2011; Moran-Crusio

TET2	P1367S	MUT	MDS/MPN	i sar. 2011)
TET2	F1368L	MUT	AML	(Jankowska i sar. 2009)
TET2	F1368Y	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	G1370V	MUT	AML	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	T1372I	MUT	MPN	(Nibourel i sar. 2010; Forbes i sar. 2011)
TET2	D1376E	MUT	AML	(Schaub i sar. 2010; Forbes i sar. 2011)
TET2	D1376G	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	D1376N	MUT	MPN	(Rocquain i sar. 2010)
TET2	H1380Y	MUT	MDS, MDS/MPN, AML	(Forbes i sar. 2011; Olcaydu i sar. 2011)
TET2	H1382Y	MUT	AML	(Smith i sar. 2010; Chou i sar. 2011; Forbes i sar. 2011; Metzeler i sar. 2011; Gaidzik i sar. 2012; Nomdedéu i sar. 2012; Weissmann i sar. 2012)
TET2	D1384H	MUT	MDS, AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	L1398R	MUT	MDS	(Rocquain i sar. 2010; Kosmider i sar. 2011)
TET2	V1417F	MUT	MDS, MDS/MPN, AML, limfoidni maligniteti	(Langemeijer i sar. 2009; Forbes i sar. 2011)
TET2	P1419R	MUT	MDS, AML, limfoidni maligniteti	(Abdel-Wahab i sar. 2009; Forbes i sar. 2011; Metzeler i sar. 2011; Quivoron i sar. 2011; Sekeres i sar. 2012; Sugimoto i sar. 2012)
TET2	D1427Y	MUT	AML	(Rocquain i sar. 2010; Forbes i sar. 2011; Quivoron i sar. 2011; Weissmann i sar. 2012)
TET2	R1440Q	SNP		(Chou i sar. 2011; Forbes i sar. 2011)
TET2	A1505T	SNP		(Olcaydu i sar. 2011)
TET2	A1505V	MUT	AML	(Metzeler i sar. 2011)
TET2	E1513G	SNP		(Forbes i sar. 2011)
TET2	V1521L	SNP		(Chou i sar. 2011)
TET2	S1556Y	MUT	AML	(Ottone i sar. 2012)
TET2	Y1579S	MUT	AML	(Makishima i sar. 2011)
TET2	N1641I	MUT	AML	(Chou i sar. 2011; Forbes i sar. 2011)
				(Abdel-Wahab i sar. 2010; Forbes i

TET2	G1697R	SNP		sar. 2011)
TET2	M1701I	SNP		(Metzeler i sar. 2011)
				(Sherry i sar. 2001; Langemeijer i sar. 2009; Smith i sar. 2010; Kohlmann i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	G1719R	MUT	MPN	(Makishima i sar. 2011)
TET2	L1721W	SNP		(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
				(Abdel-Wahab i sar. 2009; Langemeijer i sar. 2009; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Olcaydu i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	P1723S	SNP		(Abdel-Wahab i sar. 2009)
				(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	H1757D	MUT	MDS/MPN	(Abdel-Wahab i sar. 2009)
TET2	I1762V	SNP		(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
				(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
TET2	H1778R	SNP		(Sherry i sar. 2001; Langemeijer i sar. 2009; Nibourel i sar. 2010; Kohlmann i sar. 2011; Langemeijer i sar. 2011; Martínez-Avilés i sar. 2012; Weissmann i sar. 2012)
				(Chou i sar. 2011; Forbes i sar. 2011)
TET2	L1801F	MUT	AML	(Abdel-Wahab i sar. 2009)
TET2	C1811R	MUT	MDS/MPN	(Sherry i sar. 2001)
TET2	L1816F	SNP		(Ko i sar. 2010; Moran-Crusio i sar. 2011)
TET2	H1817N	MUT	MPN, AML	(Abdel-Wahab i sar. 2009)
TET2	Q1828L	MUT	MPN	(Martínez-Avilés i sar. 2012)
TET2	A1831S	SNP		(Metzeler i sar. 2011)
TET2	V1833I	SNP		(Forbes i sar. 2011)
TET2	A1863D	MUT	AML	(Chou i sar. 2011)
TET2	A1865G	MUT	AML	(Forbes i sar. 2011)
TET2	H1868Y	MUT	MPN, AML	(Forbes i sar. 2011; Weissmann i sar. 2012)
TET2	S1870P	MUT	AML	(Chou i sar. 2011; Forbes i sar. 2011)
TET2	I1871S	MUT	AML	(Chou i sar. 2011; Forbes i sar. 2011)

TET2	I1873N	MUT	MDS	2011)
TET2	I1873T	MUT	MDS, MPN, MDS/MPN, AML	(Smith i sar. 2010) (Abdel-Wahab i sar. 2009; Delhommeau i sar. 2009; Jankowska i sar. 2009; Kosmider i sar. 2009 (a); (b); Langemeijer i sar. 2009; Tefferi i sar. 2009 (b); Couronné i sar. 2010; Ko i sar. 2010; Kohlmann i sar. 2010; Chou i sar. 2011; Forbes i sar. 2011; Konstandin i sar. 2011; Metzeler i sar. 2011; Traina i sar. 2012; Weissmann i sar. 2012)
TET2	C1875R	MUT	MDS	(Langemeijer i sar. 2009; Forbes i sar. 2011)
TET2	A1876V	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	H1881R	MUT	MPN, MDS/MPN	(Abdel-Wahab i sar. 2009; Tefferi i sar. 2009 (a); Kohlmann i sar. 2010)
TET2	A1882P	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	T1884A	MUT	MDS/MPN, AML	(Kosmider i sar. 2009 (b); Kohlmann i sar. 2010; Gaidzik i sar. 2012; Weissmann i sar. 2012)
TET2	P1889L	MUT	AML	(Forbes i sar. 2011)
TET2	R1896M	MUT	AML	(Delhommeau i sar. 2009; Forbes i sar. 2011; Moran-Crusio i sar. 2011)
TET2	R1896S	MUT	MDS	(Langemeijer i sar. 2009; Forbes i sar. 2011)
TET2	T1898F	MUT	MPN, MDS/MPN, AML, limfoidni maligniteti	(Delhommeau i sar. 2009; Forbes i sar. 2011; Kosmider i sar. 2011; Quivoron i sar. 2011; Pérez i sar. 2012)
TET2	L1899P	MUT	AML	(Forbes i sar. 2011; Metzeler i sar. 2011)
TET2	V1900A	MUT	AML	(Abdel-Wahab i sar. 2009; Gaidzik i sar. 2012)
TET2	H1904R	MUT	MDS, MPN, AML	(Ko i sar. 2010; Smith i sar. 2010; Traina i sar. 2012; Weissmann i sar. 2012)
TET2	G1913D	MUT	MDS	(Jankowska i sar. 2009 Langemeijer i sar. 2009; Forbes i sar. 2011)
TET2	A1919V	MUT	MPN	(Abdel-Wahab i sar. 2009)
TET2	R1926H	MUT	MDS/MPN	(Abdel-Wahab i sar. 2009)

TET2	E1929K	SNP		(Sherry i sar. 2001)
TET2	P1941S	MUT	MDS/MPN, AML	(Abdel-Wahab i sar. 2009; Smith i sar. 2010)
TET2	R1966H	MUT	MDS/MPN	(Abdel-Wahab i sar. 2009)
TET2	R1974M	MUT	MDS/MPN	(Abdel-Wahab i sar. 2009)
TET2	T1979P	MUT	MDS	(Sugimoto i sar. 2012)
TET2	R2000K	MUT	AML	(Abdel-Wahab i sar. 2009; 2010; Forbes i sar. 2011)

Literatura

Abdel-Wahab O i sar. 2009. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood*. 114(1): 144-147.

Abdel-Wahab O i sar. 2010. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res*. 70(2): 447-452.

Abdel-Wahab O, Pardanani A, Patel J, Wadleigh M, Lasho T, Heguy A, Beran M, Gilliland DG, Levine RL, Tefferi A. 2011 (a). Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. *Leukemia*. 25(7): 1200-1202.

Abdel-Wahab O, Pardanani A, Rampal R, Lasho TL, Levine RL, Tefferi A. 2011 (b). DNMT3A mutational analysis in primary myelofibrosis, chronic myelomonocytic leukemia and advanced phases of myeloproliferative neoplasms. *Leukemia*. 25(7): 1219-1220.

Bacher U, Weissmann S, Kohlmann A, Schindela S, Alpermann T, Schnittger S, Kern W, Haferlach T, Haferlach C. 2012. TET2 deletions are a recurrent but rare phenomenon in myeloid malignancies and are frequently accompanied by TET2 mutations on the remaining allele. *Br J Haematol*. 156(1): 67-75.

Beer PA i sar. 2010. Two routes to leukemic transformation after a JAK2 mutation-positive myeloproliferative neoplasm. *Blood*. 115(14): 2891-2900.

Bejar R i sar. 2011. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*. 364(26): 2496-2506.

Boulwood J, Perry J, Zaman R, Fernandez-Santamaria C, Littlewood T, Kusec R, Pellagatti A, Wang L, Clark RE, Wainscoat JS. 2010. High-density single nucleotide polymorphism array analysis and ASXL1 gene mutation screening in chronic myeloid leukemia during disease progression. *Leukemia*. 24(6): 1139-1145.

Brecqueville M i sar. 2012. Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. *Genes Chromosomes Cancer*. 51(8): 743-755.

Brecqueville M, Cervera N, Gelsi-Boyer V, Murati A, Adélaïde J, Chaffanet M, Rey J, Vey N, Mozziconacci MJ, Birnbaum D. 2011. Rare mutations in DNMT3A in myeloproliferative neoplasms and myelodysplastic syndromes. *Blood Cancer J.* 1(5): e18.

Capello D i sar. 2011. Mutations of CD79A, CD79B and EZH2 genes in immunodeficiency-related non-Hodgkin lymphomas. *Br J Haematol.* 152(6): 777-780.

Chou WC i sar. 2011. TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood.* 118(14): 3803-3810.

Couronné L i sar. 2010. Analyses of TET2 mutations in post-myeloproliferative neoplasm acute myeloid leukemias. *Leukemia.* 24(1): 201-203.

Couronné L, Bastard C, Bernard OA. 2012. TET2 and DNMT3A mutations in human T-cell lymphoma. *N Engl J Med.* 366(1): 95-96.

Delhommeau F i sar. 2009. Mutation in TET2 in myeloid cancers. *N Engl J Med.* 360(22): 2289-2301.

Ernst T i sar. 2010. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet.* 42(8): 722-726.

Ewalt M, Galili NG, Mumtaz M, Churchill M, Rivera S, Borot F, Raza A, Mukherjee S. 2011. DNMT3a mutations in high-risk myelodysplastic syndrome parallel those found in acute myeloid leukemia. *Blood Cancer J.* 1(3): e9.

Flach J, Dicker F, Schnittger S, Kohlmann A, Haferlach T, Haferlach C. 2010. Mutations of JAK2 and TET2, but not CBL are detectable in a high portion of patients with refractory anemia with ring sideroblasts and thrombocytosis. *Haematologica.* 95(3): 518-519.

Forbes SA i sar. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 39(Database issue): D945-D950.

Fried I, Bodner C, Pichler MM, Lind K, Beham-Schmid C, Quehenberger F, Sperr WR, Linkesch W, Sill H, Wölfler A. 2012. Frequency, onset and clinical impact of somatic DNMT3A mutations in therapy-related and secondary acute myeloid leukemia. *Haematologica.* 97(2): 246-250.

Gaidzik VI i sar. 2012. TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. *J Clin Oncol.* 30(12): 1350-1357.

Gelsi-Boyer V i sar. 2010. ASXL1 mutation is associated with poor prognosis and acute transformation in chronic myelomonocytic leukaemia. *Br J Haematol.* 151(4): 365-375.

Grossmann V i sar. 2012. EZH2 mutations and their association with PICALM-MLLT10 positive acute leukaemia. *Br J Haematol.* 157(3): 387-390.

Grossmann V i sar. 2013. The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. *Genes Chromosomes Cancer*. 52(4): 410-422.

Grossmann V, Kohlmann A, Eder C, Haferlach C, Kern W, Cross NC, Haferlach T, Schnittger S. 2011. Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in >80% of patients with TET2 and EZH2 being of high prognostic relevance. *Leukemia*. 25(5): 877-879.

Guglielmelli P i sar. 2011. EZH2 mutational status predicts poor survival in myelofibrosis. *Blood*. 118(19): 5227-5234.

Jankowska AM i sar. 2011. Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood*. 118(14): 3932-3941.

Jankowska AM, Szpurka H, Tiu RV, Makishima H, Afable M, Huh J, O'Keefe CL, Ganetzky R, McDevitt MA, Maciejewski JP. 2009. Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood*. 113(25): 6403-6410.

Ko M i sar. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*. 468(7325): 839-843.

Kohlmann A i sar. 2010. Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. *J Clin Oncol*. 28(24): 3858-3865.

Kohlmann A i sar. 2011. The Interlaboratory ROBustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories. *Leukemia*. 25(12): 1840-1848.

Konstandin N i sar. 2011. Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia*. 25(10): 1649-1652.

Kosmider O i sar; GOELAMS Investigators. 2011. TET2 mutations in secondary acute myeloid leukemias: a French retrospective study. *Haematologica*. 96(7): 1059-1063.

Kosmider O i sar; Groupe Francophone des Myélodysplasies. 2009 (a). TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs). *Blood*. 114(15): 3285-3291.

Kosmider O i sar; Groupe Francophone des Myélodysplasies. 2009 (b). TET2 gene mutation is a frequent and adverse event in chronic myelomonocytic leukemia. *Haematologica*. 94(12): 1676-1681.

Langemeijer SM i sar. 2009. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet.* 41(7): 838-842.

Langemeijer SM i sar. 2011. TET2 mutations in childhood leukemia. *Leukemia.* 25(1): 189-192.

LaRochelle O i sar. 2011. Do AML patients with DNMT3A exon 23 mutations benefit from idarubicin as compared to daunorubicin? A single center experience. *Oncotarget.* 2(11): 850-861.

Ley TJ i sar. 2010. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med.* 363(25): 2424-2433.

Li Y, Zhang DF, Zhang SW, Zeng Y, Yao YG. 2012. Screening for mutation R882 in the DNMT3A gene in Chinese patients with hematological disease. *Int J Hematol.* 96(2): 229-233.

Lin J i sar. 2011. Recurrent DNMT3A R882 mutations in Chinese patients with acute myeloid leukemia and myelodysplastic syndrome. *PLoS One.* 6(10): e26906.

Lu Q, Chen Y, Wang H, Li Z. 2013. DNMT3A mutations and clinical features in Chinese patients with acute myeloid leukemia. *Cancer Cell Int.* 13(1): 1.

Makishima H i sar. 2010. Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. *Leukemia.* 24(10): 1799-1804.

Makishima H i sar. 2011. CBL, CBLB, TET2, ASXL1, and IDH1/2 mutations and additional chromosomal aberrations constitute molecular events in chronic myelogenous leukemia. *Blood.* 117(21): e198-e206.

Marková J i sar. 2012. Prognostic impact of DNMT3A mutations in patients with intermediate cytogenetic risk profile acute myeloid leukemia. *Eur J Haematol.* 88(2): 128-135.

Martínez-Avilés L, Besses C, Álvarez-Larrán A, Torres E, Serrano S, Bellosillo B. 2012. TET2, ASXL1, IDH1, IDH2, and c-CBL genes in JAK2- and MPL-negative myeloproliferative neoplasms. *Ann Hematol.* 91(4): 533-541.

McCabe MT i sar. 2012. Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci USA.* 109(8): 2989-2994.

Metzeler KH i sar. 2011. TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol.* 29(10): 1373-1381.

Moran-Crusio K i sar. 2011. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell.* 20(1): 11-24.

Morin RD i sar. 2010. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet.* 42(2): 181-185.

Nibourel O i sar. 2010. Incidence and prognostic value of TET2 alterations in de novo acute myeloid leukemia achieving complete remission. *Blood.* 116(7): 1132-1135.

Nikoloski G i sar. 2010. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet.* 42(8): 665-667.

Nomdedéu J i sar. 2012. Adverse impact of IDH1 and IDH2 mutations in primary AML: experience of the Spanish CETLAM group. *Leuk Res.* 36(8): 990-997.

Olcaydu D i sar. 2011. The role of the JAK2 GGCC haplotype and the TET2 gene in familial myeloproliferative neoplasms. *Haematologica.* 96(3): 367-374.

Ottone T i sar. 2012. Comparative molecular analysis of therapy-related and de novo acute promyelocytic leukemia. *Leuk Res.* 36(4): 474-478.

Pérez B i sar. 2010. Genetic typing of CBL, ASXL1, RUNX1, TET2 and JAK2 in juvenile myelomonocytic leukaemia reveals a genetic profile distinct from chronic myelomonocytic leukaemia. *Br J Haematol.* 151(5): 460-468.

Pérez C i sar. 2012. TET2 mutations are associated with specific 5-methylcytosine and 5-hydroxymethylcytosine profiles in patients with chronic myelomonocytic leukemia. *PLoS One.* 7(2): e31605.

Pezzi A i sar. 2012. DNMT3A Mutations in Patients with Acute Myeloid Leukemia in South Brazil. *Adv Hematol.* 2012: 697691.

Pronier E i sar. 2011. Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. *Blood.* 118(9): 2551-2555.

Quivoron C i sar. 2011. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell.* 20(1): 25-38.

Rao N, Butcher CM, Lewis ID, Ross DM, Melo JV, Scott HS, Bardy PG, D'Andrea RJ. 2012. Clonal and lineage analysis of somatic DNMT3A and JAK2 mutations in a chronic phase polycythemia vera patient. *Br J Haematol.* 156(2): 268-270.

Rocquain J i sar. 2010. Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC Cancer.* 10: 401.

Schaub FX, Looser R, Li S, Hao-Shen H, Lehmann T, Tichelli A, Skoda RC. 2010. Clonal analysis of TET2 and JAK2 mutations suggests that TET2 can be a late event in the progression of myeloproliferative neoplasms. *Blood.* 115(10): 2003-2007.

Schnittger S i sar. 2013. ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. *Leukemia*. 27(1): 82-91.

Sekeres MA i sar. 2012. Phase 2 study of the lenalidomide and azacitidine combination in patients with higher-risk myelodysplastic syndromes. *Blood*. 120(25): 4945-4951.

Shen Y i sar. 2011. Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood*. 118(20): 5593-5603.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 29(1): 308-311.

Shih AH, Chung SS, Dolezal EK, Zhang SJ, Abdel-Wahab OI, Park CY, Nimer SD, Levine RL, Klimek VM. 2013. Mutational analysis of therapy-related myelodysplastic syndromes and acute myelogenous leukemia. *Haematologica*. 98(6): 908-912.

Simon C i sar. 2012. A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev*. 26(7): 651-656.

Smith AE i sar. 2010. Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value. *Blood*. 116(19): 3923-3932.

Stegelmann F, Bullinger L, Schlenk RF, Paschka P, Griesshammer M, Blersch C, Kuhn S, Schauer S, Döhner H, Döhner K. 2011. DNMT3A mutations in myeloproliferative neoplasms. *Leukemia*. 25(7): 1217-1219.

Sugimoto Y i sar. 2010. Spectrum of molecular defects in juvenile myelomonocytic leukaemia includes ASXL1 mutations. *Br J Haematol*. 150(1): 83-87.

Sugimoto Y i sar. 2012. Cytogenetic and molecular predictors of response in patients with myeloid malignancies without del[5q] treated with lenalidomide. *J Hematol Oncol*. 5: 4.

Tefferi A i sar. 2009 (a). Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFR α correlates. *Leukemia*. 23(5): 900-904.

Tefferi A i sar. 2009 (b). TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. *Leukemia*. 23(5): 905-911.

Thiel A i sar. 2011. Comprehensive array CGH of normal karyotype myelodysplastic syndromes reveals hidden recurrent and individual genomic copy number alterations with prognostic relevance. *Leukemia*. 25(3): 387-399.

Thol F i sar. 2011 (a). Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J Clin Oncol*. 29(21): 2889-2896.

Thol F, Heuser M, Damm F, Klusmann JH, Reinhardt K, Reinhardt D. 2011 (b). DNMT3A mutations are rare in childhood acute myeloid leukemia. *Haematologica*. 96(8): 1238-1240.

Thol F, Winschel C, Lüdeking A, Yun H, Friesen I, Damm F, Wagner K, Krauter J, Heuser M, Ganser A. 2011 (c). Rare occurrence of DNMT3A mutations in myelodysplastic syndromes. *Haematologica*. 96(12): 1870-1873.

Traina F i sar. 2012. Single nucleotide polymorphism array lesions, TET2, DNMT3A, ASXL1 and CBL mutations are present in systemic mastocytosis. *PLoS One*. 7(8): e43090.

Voso MT i sar. 2011. Role of BCL2L10 methylation and TET2 mutations in higher risk myelodysplastic syndromes treated with 5-azacytidine. *Leukemia*. 25(12): 1910-1913.

Walter MJ i sar. 2011. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*. 25(7): 1153-1158.

Wang J i sar. 2013. TET2, ASXL1 and EZH2 mutations in Chinese with myelodysplastic syndromes. *Leuk Res*. 37(3): 305-311.

Weissmann S i sar. 2012. Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia*. 26(5): 934-942.

Wlodarski MW, Mötter J, Gorr TA, Olk-Batz C, Hasle H, Dworzak M, Niemeyer CM, Flotho C. 2011. Abnormal promoter DNA methylation in juvenile myelomonocytic leukemia is not caused by mutation in DNMT3A. *Blood*. 118(16): 4490-4491.

Yamashita Y i sar. 2010. Array-based genomic resequencing of human leukemia. *Oncogene*. 29(25): 3723-3731.

Yan XJ i sar. 2011. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet*. 43(4): 309-315.

Yoshida K i sar. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 478(7367): 64-69.

Spisak slika i tabela

Slike

- Slika 1. Genetičke promene, koje dovode do patogeneze maligniteta. Prikazane su sukcesivne mitotičke deobe od oplođene jajne ćelije do maligne ćelije, sa naznačenim vremenom sticanja somatskih mutacija i procesima koji tome doprinose. Modifikovano prema Stratton i sar. 2009.
- Slika 2. Klasifikacija mijeloidnih maligniteta. Modifikovano prema Tefferi i sar. 2011.
- Slika 3. Shematski prikaz pet klasa gena uključenih u patogenezu mijeloidnih maligniteta. Modifikovano prema Murati i sar. 2012.
- Slika 4. Shematski prikaz primene modela farmakofora. Modifikovano prema Yang SY 2010.
- Slika 5. Shematski prikaz algoritma metode informacionog spektra – ISM (na primeru WT1 i EZH2 proteina).
- Slika 6. Tri primera ROC krivih. PRED1 predstavlja primer performanse idealnog alata, sa $AUC = 1$; PRED2 predstavlja primer performanse alata koji daje nasumične predikcije, sa $AUC = 0,500$; PRED3 predstavlja primer performanse realnog alata, sa $AUC = 0,701$.
- Slika 7. Integrativni algoritam za predikciju protein-protein interakcija.
- Slika 8. Informacioni spektri četiri glavne izoforme humanog WT1. Dominantan je lokalni maksimum na frekvenci $F(0,00195)$.

- Slika 9. Informacioni spektri sekvenci WT1 miša, pacova i svinje. Dominantan je lokalni maksimum na frekvenci $F(0,00195)$.
- Slika 10. Predviđeni novi interaktori WT1, koeksprimirani sa WT1 u K562 ćelijskoj liniji, i njihove međusobne veze; debljina linija zavisi od STRING skora poverenja.
- Slika 11. Spektralna analiza proteina NISCH. A. Spektar proteina NISCH, sa označenim karakterističnim frekvencama, $F(0,152)$ i $F(0,426)$. B. Domen koji najviše doprinosi intenzitetu amplitude na frekvenci $F(0,152)$. C. Domen koji najviše doprinosi intenzitetu amplitude na frekvenci $F(0,426)$.
- Slika 12. Vizualizacija GO termina, koji reprezentuju biološke procese u koje je uključen domen 680-739 proteina NISCH (asociran sa $F(0,152)$).
- Slika 13. Vizualizacija GO termina, koji reprezentuju biološke procese u koje je uključen domen 966-1020 proteina NISCH (asociran sa $F(0,426)$).
- Slika 14. Distribucija mutacija u mijeloidnim malignitetima (crveno) i SNP-ova (plavo) duž sekvenci ASXL1, EZH2, DNMT3A i TET2 epigenetičkih regulatora; crnim su označeni konzervirani funkcionalni domeni.
- Slika 15. Poređenje tačnosti, senzitivnosti i specifičnosti alata PolyPhen-2 i SIFT u predviđanju funkcionalnih efekata aminokiselinskih supstitucija iz SVE_AAS (zeleni stubići) i nCFD_AAS (ljubičasti stubići) skupova.
- Slika 16. ISM algoritam za predikciju funkcionalnih efekata aminokiselinskih supstitucija.
- Slika 17. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu ASXL1. A. Informacioni spektar divljeg tipa proteina ASXL1. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala

skeniranje na frekvenci $F(0,476)$. D. Raspodela ISM skorova za SNP-ove i mutacije u ASXL1.

Slika 18. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu EZH2. A. Informacioni spektar divljeg tipa proteina EZH2. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala skeniranje na frekvenci $F(0,411)$. D. Raspodela ISM skorova za SNP-ove i mutacije u EZH2.

Slika 19. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu DNMT3A. A. Informacioni spektar divljeg tipa proteina DNMT3A. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala skeniranje na frekvenci $F(0,071)$. D. Raspodela ISM skorova za SNP-ove i mutacije u DNMT3A.

Slika 20. Predikcija funkcionalnih efekata aminokiselinskih supstitucija u proteinu TET2. A. Informacioni spektar divljeg tipa proteina TET2. B. Selekcija značajne frekvence pomoću Mann-Whitney U testa. C. Virtuelno Ala skeniranje na frekvenci $F(0,491)$. D. Raspodela ISM skorova za SNP-ove i mutacije u TET2.

Slika 21. ROC analiza PolyPhen-2, SIFT i ISM skorova kao prediktora funkcionalnih efekata aminokiselinskih supstitucija pozicioniranih u nCFD.

Slika 22. ROC analiza PolyPhen-2, SIFT i ISM predikcija funkcionalnih efekata aminokiselinskih supstitucija pozicioniranih u nCFD.

Tabele

Tabela 1. Vrednosti potencijala elektron jon interakcija (EIIP) za 20 aminokiselina.

- Tabela 2. Identifikatori 12 studija iz *Gene Expression Omnibus* baze, sa 41 uzorkom K562 ćelija u bazičnom, netretiranom stanju.
- Tabela 3. Primer unakrsne tabele, sa tačno pozitivnim, tačno negativnim, lažno pozitivnim i lažno negativnim predikcijama alata za predviđanje funkcionalnih efekata aminokiselinskih supstitucija.
- Tabela 4. Proteinski partneri WT1 identifikovani do kraja 2011. godine.
- Tabela 5. S/N(0,00195) iz kros-spektara WT1 i njegovih poznatih interaktora, u kojima je F(0,00195) dominantna frekvenca.
- Tabela 6. DAVID *enrichment* skorovi za funkcionalne klastere dobijene analizom skupa proteina LISTA1.
- Tabela 7. LISTA3 dobijena STRING analizom WT1 i njegovih poznatih interaktora.
- Tabela 8. LISTA4 predstavlja skup kandidata za nove interaktore WT1. Prikazana je koekspresija proteina iz ovog skupa sa WT1 u K562 ćelijskoj liniji; koeksprimiranim se smatraju proteini prisutni (P+M) u $\geq 80\%$ uzoraka.
- Tabela 9. GO termini koji reprezentuju biološke procese u koje je uključen domen 680-739 proteina NISCH (asociran sa F(0,152)); *enrichment* na svim terminima je statistički značajno ($p \leq 0,01$).
- Tabela 10. GO termini koji reprezentuju biološke procese u koje je uključen domen 966-1020 proteina NISCH (asociran sa F(0,426)); *enrichment* na svim terminima je statistički značajno ($p \leq 0,01$).
- Tabela 11. Pregled SVE_AAS, skupa svih aminokiselinskih supstitucija u epigenetičkim regulatorima ASXL1, EZH2, DNMT3A i TET2.

- Tabela 12. Epigenetički regulatori ASXL1, EZH2, DNMT3A i TET2: njihovi ID u UniProt bazi, konzervirani funkcionalni domeni i odgovarajuća literatura.
- Tabela 13. Statistički parametri efikasnosti predviđanja funkcionalnih efekata aminokiselinskih supstitucija PolyPhen-2 i SIFT na nCFD_AAS skupu (p vrednosti su dobijene Fišerovim testom).
- Tabela 14. Spirmanove korelacije CFD/nCFD pozicije aminokiselinskih supstitucija i PolyPhen-2 i SIFT predikcija; sve korelacije su statistički značajne ($p < 0,001$).
- Tabela 15. Statistički parametri efikasnosti predviđanja funkcionalnih efekata aminokiselinskih supstitucija ISM algoritma na nCFD-AAS skupu (p vrednosti su dobijene Fišerovim testom).

List of Figures and Tables

Figures

- Figure 1. Genetic changes leading to pathogenesis of malignancies. Successive mitotic divisions, from the fertilized egg to the malignant cell, are shown, with the marked time of acquiring somatic mutations and the processes contributing to that. Modified from Stratton et al. 2009.
- Figure 2. Classification of myeloid malignancies. Modified from Tefferi et al. 2011.
- Figure 3. Schematic summary of five classes of genes involved in pathogenesis of myeloid malignancies. Modified from Murati et al. 2012.
- Figure 4. Schematic summary of the pharmacophore model application. Modified from Yanf SY 2010.
- Figure 5. Scheme of the algorithm based on the Informational Spectrum Method – ISM (example of WT1 and EZH2 proteins).
- Figure 6. Three examples of ROC curves. PRED1 represents the performance of a perfect tool, with $AUC = 1$; PRED2 represents the performance of a tool resulting in random predictions, with $AUC = 0.500$; PRED3 represents performance of a real tool, with $AUC = 0.701$.
- Figure 7. Integrative algorithm for protein-protein interactions prediction.
- Figure 8. Informational spectra of four major isoforms of human WT1. There is a dominant peak at frequency $F(0.00195)$.
- Figure 9. Informational spectra of WT1 sequences of mouse, rat and pig. There is a dominant peak at frequency $F(0.00195)$.

- Figure 10. Predicted new interactors of WT1, co-expressed with WT1 in K562 cell line, and their links; weight of the lines depends on the STRING confidence score.
- Figure 11. Spectral analysis of NISCH protein. A. Spectrum of NISCH protein, with marked characteristic frequencies, $F(0.152)$ and $F(0.426)$. B. Domain attributing the most to the amplitude intensity at the frequency $F(0.152)$. C. Domain attributing the most to the amplitude intensity at the frequency $F(0.426)$.
- Figure 12. Visualization of GO terms, representing biological processes in which the NISCH domain 680-739 (associated with $F(0.152)$) is involved.
- Figure 13. Visualization of GO terms, representing biological processes in which the NISCH domain 966-1020 (associated with $F(0.426)$) is involved.
- Figure 14. Distribution of mutations in myeloid malignancies (red) and SNPs (blue) along the sequences of epigenetic regulators ASXL1, EZH2, DNMT3A and TET2; black boxes are conserved functional domains.
- Figure 15. Comparison of accuracy, sensitivity and specificity for PolyPhen-2 and SIFT in predicting functional effects of amino acid substitutions between SCE_AAS (green bars) and nCFD_AAS (violet bars) sets.
- Figure 16. ISM algorithm for prediction of functional effects of amino acid substitutions.
- Figure 17. Predicting functional effects of amino acid substitutions in ASXL1 protein. A. Informational spectrum of wild type ASXL1 protein. B. Selection of significant frequency using Mann-Whitney U test. C. Virtual Ala screening at the frequency $F(0.476)$. D. Distribution of ISM scores for SNPs and mutations in ASXL1.

Figure 18. Predicting functional effects of amino acid substitutions in EZH2 protein. A. Informational spectrum of wild type EZH2 protein. B. Selection of significant frequency using Mann-Whitney U test. C. Virtual Ala screening at the frequency F(0.411). D. Distribution of ISM scores for SNPs and mutations in EZH2.

Figure 19. Predicting functional effects of amino acid substitutions in DNMT3A protein. A. Informational spectrum of wild type DNMT3A protein. B. Selection of significant frequency using Mann-Whitney U test. C. Virtual Ala screening at the frequency F(0.071). D. Distribution of ISM scores for SNPs and mutations in DNMT3A.

Figure 20. Predicting functional effects of amino acid substitutions in TET2 protein. A. Informational spectrum of wild type TET2 protein. B. Selection of significant frequency using Mann-Whitney U test. C. Virtual Ala screening at the frequency F(0.491). D. Distribution of ISM scores for SNPs and mutations in TET2.

Figure 21. ROC analysis of PolyPhen-2, SIFT and ISM scores as predictors of functional effects of amino acid substitutions positioned in the nCFDs.

Figure 22. ROC analysis of PolyPhen-2, SIFT and ISM predictions of functional effects of amino acid substitutions positioned in the nCFDs.

Tables

Table 1. Values of Electron-Ion Interactions Potential (EIIP) for 20 amino acids.

Table 2. Identifiers of 12 studies from Gene Expression Omnibus database, with 41 cases of K562 cells in basic, untreated state.

- Table 3. Example of the crosstabs, with true positive, true negative, false positive and false negative predictions of a tool for functional effects of amino acid substitutions prediction.
- Table 4. Protein partners of WT1, identified by the end of 2011.
- Table 5. S/N(0.00195) from the cross-spectra of WT1 and its known interactors, in which the F(0.00195) is a dominant frequency.
- Table 6. DAVID enrichment scores for the functional clusters resulting from the analysis of the LISTA1 set of proteins.
- Table 7. LISTA3, resulting from the STRING analysis of WT1 and its known interactors.
- Table 8. LISTA4, representing a set of candidates for new WT1 interactors. Co-expression of the proteins from this set with WT1 in K562 cell line is shown; proteins are considered co-expressed if present (P+M) in $\geq 80\%$ of cases.
- Table 9. GO terms representing biological processes in which the NISCH domain 680-739 (associated with F(0.152)) is involved; enrichment of all of the terms is statistically significant ($p \leq 0.01$).
- Table 10. GO terms representing biological processes in which the NISCH domain 966-1020 (associated with F(0.426)) is involved; enrichment of all of the terms is statistically significant ($p \leq 0.01$).
- Table 11. Summary of SVE_AAS, a set containing all amino acid substitutions in epigenetic regulators ASXL1, EZH2, DNMT3A and TET2.
- Table 12. Epigenetic regulators ASXL1, EZH2, DNMT3A and TET2: their UniProt IDs, conserved functional domains and related literature.

- Table 13. Statistical parameters of efficacy of PolyPhen-2 and SIFT in predicting functional effects of amino acid substitutions in the nCFD_AAS set (p values are from the Fisher's test).
- Table 14. Spearman correlations of CFD/nCFD positions of amino acid substitutions and PolyPhen-2 and SIFT predictions; all correlations are statistically significant ($p < 0.001$).
- Table 15. Statistical parameters of efficacy of ISM algorithm in predicting functional effects of amino acid substitutions in the nCFD_AAS set (p values are from the Fisher's test).

Biografija autora

Branislava Gemović je rođena 30.03.1982. u Zemunu. Biološki fakultet Univerziteta u Beogradu, studijska grupa Biologija, upisala je školske 2001/2002. godine. Diplomirala je 2008. godine, sa prosečnom ocenom 9,03 i ocenom 10 na diplomskom radu. Školske 2009/2010. je upisala doktorske studije na Biološkom fakultetu Univerziteta u Beogradu, smer Molekularna biologija.

Od 2009. godine, zaposlena je kao istraživač pripravnik u Centru za multidisciplinarna istraživanja i inženjering Instituta za nuklearne nauke Vinča Univerziteta u Beogradu, a u zvanje istraživač saradnik izabrana je 2012. godine. Trenutno je angažovana na projektu Ministarstva prosvete, nauke i tehnološkog razvoja ON173001: „Primena EIIP/ISM bioinformatičke platforme u otkrivanju novih terapijskih targeta i potencijalnih terapijskih molekula“.

Od 2009. do 2013. je učestvovala u COST akciji BM0801 „EuGESMA: European Genomics and Epigenomics Study on MDS and AML“. Pohađala je dve bioinformatičke škole: 2010. godine – analiza *microarray* podataka (Barselona, Španija) i 2012. godine – analiza sekvenciranja RNK (Modena, Italija). Dobitnik je granta za istraživanje „Analysis of Wilms’ Tumor 1 (WT1) mutations in Acute Myeloid Leukemia (AML)“, koje je izvršeno u Center for Hemato-Oncology, Barts Cancer Institute, Queen Mary University of London, pod rukovodstvom dr Jude Fitzgibbon.

Aktivna je u promociji nauke kroz učešće u Otvorenim vratima Vinče i Vinčinoj letnjoj sceni. Od 2013. godine je koordinator radionica Centra za multidisciplinarna istraživanja i inženjering u okviru projekta Vinčinih naučionica.

Do sada je bila autor/koautor 4 naučna rada u časopisima međunarodnog značaja i 5 saopštenja sa naučnih skupova od međunarodnog značaja štampanih u izvodu. Prema Science Citation Index-u, citirana je 5 puta.

Прилог 1.

Изјава о ауторству

Потписани-а Бранислава Гемовић
број индекса М3003/09

Изјављујем

да је докторска дисертација под насловом

Биоинформатичка анализа протеина укључених у патогенезу мијелоидних
малигнитета

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 01.12.2014.

Гемовић Б.

Прилог 2.

**Изјава о истоветности штампане и електронске верзије
докторског рада**

Име и презиме аутора Бранислава Гемовић

Број индекса М3003/09

Студијски програм Молекуларна биологија

Наслов рада Биоинформатичка анализа протеина укључених у патогенезу
мијелоидних малигнитета

Ментори др Душанка Савић Павићевић, ванредни професор Биолошког
факултета Универзитета у Београду

Др Невена Вељковић, научни саветник Института за нуклеарне науке
Винча Универзитета у Београду

Потписани/а Бранислава Гемовић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 01.12.2014.

Ђеровић Б.

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Биоинформатичка анализа протеина укључених у патогенезу
мијелоидних малигнитета

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство – некомерцијално – без прераде
4. Ауторство – некомерцијално – делити под истим условима
5. Ауторство – без прераде
6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 01.12.2014.

Jerobalt S.

1. Ауторство - Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.
2. Ауторство – некомерцијално. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.
3. Ауторство - некомерцијално – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.
4. Ауторство - некомерцијално – делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.
5. Ауторство – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.
6. Ауторство - делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.